

A week in the life of the human brain: stable states punctuated by chaotic transitions

Maxwell Wang (✉ mbwang@cmu.edu)

Carnegie Mellon University

Max G'Sell

Carnegie Mellon University

James F. Castellano

University of Pittsburgh

R. Mark Richardson

University of Pittsburgh <https://orcid.org/0000-0003-2620-7387>

Avniel Ghuman (✉ ghumana@upmc.edu)

University of Pittsburgh

Biological Sciences - Article

Keywords:

Posted Date: October 2nd, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2752903/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

A week in the life of the human brain: stable states punctuated by chaotic transitions

Authors: Maxwell B. Wang^{*,1,2,3,4,9}, Max G'Sell^{2,5}, James F. Castellano⁶, R. Mark Richardson^{3,7,8}, and Avniel Singh Ghuman^{*,1,3,9,10}

Affiliations:

¹Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA

²Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

³Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA, USA

⁴Medical Scientist Training Program, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA

⁵Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA, USA

⁶Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA

⁷Department of Neurosurgery, Massachusetts General Hospital, Boston, MA, USA

⁸Harvard Medical School, Boston, MA, USA

⁹Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA

¹⁰Center for Neuroscience at the University of Pittsburgh, Pittsburgh, PA, USA

*Corresponding authors. Email: mbwang@alumni.cmu.edu; ghumana@upmc.edu

Abstract:

Many important neurocognitive states, such as ones related to performing natural activities and fluctuations of arousal, shift over minutes-to-days in the real-world. We analyzed 3-12 days of continuous intracranial recordings in twenty participants that freely socialized, used digital devices, slept, etc. to understand how neural dynamics form and change with behavior. Brain networks formed stable states that were predictive of both behavior and physiology. Behavior changes were associated with bursts of rapid neural fluctuations where brain networks chaotically explored many configurations before settling into new states. These trajectories traversed an hourglass-like structure, with awake and sleep at opposite ends, and an attractor state represented by default mode network activation in between. These findings illustrate ways our brains balance stability and flexibility to produce real-world behavior.

One-Sentence Summary: During a week of intracranial recordings, the brain forms a punctuated equilibrium of stable states and chaotic transitions.

Main Text:

Whether we are fatigued from attending a teleconference call or eager to read a book, whether we feel vibrant and ready to start our day or weary and winding down to sleep, many neurocognitive processes in our lives slowly fluctuate over minutes to hours. Yet, most of our understanding of human brain activity comes either from well-controlled experiments, studying reactions to carefully chosen stimuli over milliseconds to seconds, or examining spontaneous neural activity from subjects “resting” inside a neuroimaging machine. A few studies have analyzed brain state dynamics over minutes in a single sitting or repeatedly sampled a few minutes per day spread out over days to months using functional neuroimaging (1–8). A small number of studies took similar snapshot approaches in real-world settings, such as tracking depression phenotypes (9, 10) or classifying small windows of a few specific behaviors (11, 12).

As a result, it is still unclear how our brain activity continuously changes over timescales of minutes-to-hours-to-days, particularly during natural behaviors that depart from the boundaries of conventional experimental paradigms. Here we asked several questions regarding long timescale brain dynamics: What neural states emerge during natural behavior when recorded continuously over a week using intracranial monitoring? How does the brain transition between these states? Do these states and transitions follow a consistent organization with respect to behavior, physiology, and anatomy across people?

We used intracranial neural recordings (80-126 electrodes implanted per participant) in twenty neurosurgical participants undergoing evaluation for epilepsy surgery for between 75 to 283 hours (near-continuous recordings across approximately 3-12 days). During this time, participants were confined to the hospital but would freely socialize with friends, family, and staff, interact with digital devices, sleep, watch TV, and perform other volitional natural behaviors while under simultaneous neural and video monitoring. We started by examining the basic dynamics of different areas of the brain, finding that they would change in consistent manners, albeit oftentimes in a complex and nonlinear fashion. To probe this nonlinearity, we used self-supervised deep recurrent neural networks and Koopman operators to learn a “state-space” of the brain’s dynamics – a geometric representation of the brain’s activity where changing patterns of brain network activations are reflected as movement in this state-space. This state-space allows us to query brain networks’ overall organization, relationships to behavior, and primary dynamical driving forces.

Functional parcels showed consistent fluctuations over days and their dynamics displayed consistent anatomic trends.

After removing an hour before and after ictal (seizure) events as determined by the clinical team, we calculated the coherence between all pairs of electrodes in each participant every five seconds over five frequencies: theta (θ : 4-8Hz), alpha (α : 8-12Hz), low beta (β_l : 14-20Hz), high beta (β_h : 20-30Hz), and gamma (γ : 30-70Hz) (13). Electrodes were parcellated into tightly connected, anatomically compact groups of electrodes, each a “parcel” of the brain (Figure 1A). After removing parcels and activity associated to the participants’ seizure onset zones, we plotted the coherence of each parcel over the week (Figure 1B-top). Fluctuations of these coherences showed characteristic temporal scales that were repeated over different hours and days of data. We quantified this stability by how slowly each parcel’s autocorrelation decayed (timescale). Timescale differences between parcels were stable over time, indicating that parcels that fluctuated faster or slower would remain so throughout the week (Supplementary Figure S3).

Differences between parcels were quantified over all twenty participants by grouping parcels according to which of six canonical fMRI networks they fell in (“default mode”, “dorsal attention”, “salience”, “somatomotor”, “control”, and “visual” as defined in (14)). Parcels in the default mode consistently showed higher autocorrelation magnitude and longer decay timescales across our participants, whereas parcels of the salience network showed shorter timescales (Figure 1B-bottom). These findings demonstrate a temporal hierarchy separating “fast” and “slow” regions of the brain. A temporal hierarchy, typically measured using autocorrelation, has been hypothesized in the brain with transmodal systems, such as the default mode network, slowly integrating data from faster unimodal regions over seconds to minutes depending on the task (1, 2, 15, 16). Our results extend these findings to minutes-to-hours in a real-world setting during natural behavior.

Neural dynamics predicted physiology.

To assess the neurophysiological relevance of the above dynamics, we linked them to fluctuations in circadian rhythms and arousal. After grouping parcels into networks to reduce dimensionality in a data-driven fashion, we took the first half of the week for each participant and used canonical correlation analysis (CCA) (17) to identify networks that maximized correlation to time of day. We tested this group of networks on the second half of the week using permutation testing and found that 11 of the 20 participants had networks significantly linked with circadian rhythm (Figure 2A). Notably, six of the nine participants whose data lacked significant correlation to time of day had sleep disturbances such as nocturnal-awakening seizures or intentional clinical sleep deprivation, suggesting these participants had disrupted circadian rhythms.

Seven participants had sufficiently clean electrocardiogram (EKG) signals to track heart rate. Heart rate is strongly correlated with the degree of arousal (18) and is used here as a proxy for it. We used L1-regularized (19) regression over the first half of the week to identify a group of networks that predicted heart rate and tested this group on the remaining half (Figure 2B). Six of the seven participants had networks that were significantly associated with heart rate.

Brain networks underwent bursts of rapid transitions that coincided with natural behavior shifts.

We next assessed overall brain network dynamics and the relationship between these dynamics and natural behavior. The top of Figure 3A shows the brain network activation patterns from one participant for the full week and the mid/bottom of Figure 3A shows the “speed” of one participant’s brain throughout the week: how quickly the brain changed its network pattern between consecutive time windows. Times of high speed occurred in “bursts” where the brain rapidly modulated its networks before stabilizing into a new configuration (examples in Figure 3C; Figure S7 quantifies the “burstiness” of the dynamics across all participants). These results illustrate that brain network dynamics follow a punctuated equilibrium of relatively stable “equilibrium” periods of slowly changing brain networks that are “punctuated” by transitory bursts of rapidly changing brain networks.

To assess how these neural transitions related to behavioral transitions, in nine participants with high-quality video recordings, we marked periods of time when participants underwent three broad categories of behavior: interacting with a digital screen, socializing with another person, or physically manipulating an object. We marked times when participants began or ended one of these three behaviors and found that neural and behavioral transitions tended to

coincide with one another (Figure 3B; detailed examination of the relationship between the behavioral and neural states themselves is performed in a later section below).

Brain networks transitions were circuitous, unpredictable, and chaotic.

How does the brain transition between the starting and ending states of these bursts? Do these bursts take consistent paths? To answer these questions, we defined a neural state-space: a representation of the brain in a high-dimensional Euclidean space where the axes represent the activation of different brain networks. A single time window forms a point in this space where the point's position along each axis marks the coherence of each brain network during that time window. A transitory burst (a series of consecutive time windows where brain networks start in one configuration and ends in another) becomes a trajectory in this state-space: a series of points leading from a starting state to an ending state.

We found that these bursts were circuitous by measuring the total distance traversed by a trajectory (the trajectory's length) and comparing it to the trajectory's displacement (the straight-line distance between the starting and ending point). The distance was on average 8.9 times longer than the displacement across participants during transitions and the ratio for transitions is greater than during stable states (8.9 versus 6.0; Supplemental Figure S13). Examples of these circuitous trajectories from one participant are found in Figure 3C.

After grouping bursts that shared highly similar starting states, we calculated the distance between their trajectories as a function of what percentage of the trajectory was complete. If neural bursts went from their starting to ending points in a consistent manner, bursts sharing the same starting and ending points would take highly similar paths. Instead, nearly the first half of these paths displayed almost as much variability between bursts sharing a destination as between bursts with different destinations despite sharing their starting location (Figure 3D). Only when about 75% of the trajectory was complete did the variability between bursts sharing both a starting and ending point diverge from the variability of bursts sharing only a starting point by a Cohen's *d* effect size of one. Using 0-1 chaos tests, we found that chaoticity within the brain's dynamics rose during times of these transitory bursts (Figure 3E), indicating that these transitory bursts were both non-repeated and chaotic-like. Additionally, the size of these transitions and the time between them followed power laws that are oftentimes associated with chaotic and critical systems (Figure 3F) (20).

During natural behavior, instead of taking a direct, consistent route between neural states, the brain undergoes a chaotic, exploratory-like phase where the variance of its trajectories dramatically rises before stabilizing onto a destination. Upon reaching these destinations, the brain would enter stable states of decreased variability and chaoticity, presumably exploiting currently active networks to accomplish some goal until the participant's behavior changed once again. These transitions may reflect real-world correlates of task switching typically studied in lab experiments (21, 22) thought to relate to "cognitive flexibility." These dynamics extend early studies of chaos in the brain and ongoing theoretical models (23–25) based on task data in controlled settings.

Neural dynamics are driven by a central homeostatic-like attractor at the default mode.

While the brain's movement appears chaotic, is there a consistent anatomical trend and organization? Do dynamics organize consistently with respect to behavior? System dynamics, in and out of biology, are traditionally defined around their critical points: points in the system that draw in or push out the system's dynamics (26). If we describe how a ball rolls throughout hills

and valleys, we describe how it rolls away from the top of hills and towards the bottom of valleys. In metabolic physiology, we describe sodium or glucose levels by how they fluctuate relative to the homeostatic equilibrium points they stabilize to.

In order to capture the brain's complex and chaotic dynamics in a simple dynamical form, we used self-supervised deep recurrent neural networks and Koopman operators (27). We started by taking all the data from the week except for two days and learning the underlying building blocks of these dynamics. Neurocognitive states form out of combinations of these individual building blocks and their dynamics unfold according to their interactions and trends, like how words form out of combinations of letters and a sentence's meaning unfolds from the grammatical interactions of its words and phrases. More specifically, we took the original state-space and mapped each point in it onto a new nonlinear manifold where each axis of that manifold represents a single building block and the brain's overall neurocognitive state becomes a sum of these blocks. We defined this nonlinear axis transformation such that the temporal evolution of these blocks can be captured using easily interpretable linear methods, allowing us to identify its underlying dynamical drivers.

To validate our model, we took the two days that were not used to learn our neural networks and Koopman operators and annotated the participant's behavior during these days into the three major categories used in Figure 3B: watching a digital device, socializing, and physically manipulating an object. We trained linear behavioral prediction classifiers on the brain's position along the learned manifold on one day and tested them on the other day. If learning these building blocks increases classification accuracy, then the building blocks and states learned by the algorithm are neurocognitively relevant.

Learning this manifold increased our capability to predict all three natural behaviors (Figure 4B-left, $p=0.012$). This indicates that a) the manifold contained neurocognitively relevant information that could be decoded by interpretable linear operators and b) natural behavior organized consistently within this manifold. To assess parts of this organization, we asked what brain networks were associated with areas of the manifold tied to each behavior. Social interactions activated dorsal attention and somatomotor networks. Physically manipulating an object activated the dorsal attention, somatomotor, and salience networks (Figure 4B-right). Watching a digital device did not consistently activate any of the brain networks we examined across participants.

We next asked whether there was a consistent dynamical organization for how the brain moved around on this manifold. Using eigendecomposition on the Koopman operator, we found that the brain's dynamics trended towards a central attractor in every participant. This attractor is visualized in Figure 4C where we plot the flow diagram of the brain's dynamics: how the brain's state tends to change as a function of which networks are active/inactive, and their overall tendency is to drift towards a central state. This is quantified by the eigenvalues of the Koopman operator in Supplementary Figure S11, which show that in all twenty participants the brain tended to move towards this attractor. At this attractor state, the brain consistently activated the default mode network while trending towards deactivating the visual network across our participants (Figure 4D). Together, these results demonstrate that neural states organized consistently across participants with respect to both outwardly observable behavior and inwardly observable dynamical trends.

Neurocognitive states form an hourglass-like shape where the default mode attractor separates waking and sleep.

If there is a consistent behavioral and dynamical organization of the brain's nonlinear dynamics, what does this organization look like? How is the representation of different neurocognitive states geometrically organized in this space? In Figure 5A-left, for two participants, we plot a full day of the brain's trajectory in the Koopman space colored by what behavioral state the participant was in, along with the attractor state. Continuing the analogy from the prior section, this plot is an illustration of how the building blocks of human brain network activity are combined to form the words and sentences of the brain's language that underpin neurocognitive states participants went through over a day of natural behavior. Qualitatively, these day-long trajectories formed "hourglasses" where different waking behaviors formed separate quadrants in the top of the hourglass, sleep formed the bottom of the hourglass, and periods where the participant is awake but not doing any of the three annotated behaviors formed the middle funnel around the attractor state. We denote this middle funnel associated state as "wakeful rest," because during these times, participants were awake but not outwardly interacting with their environment.

We verified this structure quantitatively across participants and found that the brain's state departs further away from the central attractor during times of active behavior than times of wakeful rest (Figure 5A). We defined an axis between the center of brain states associated with sleeping and those associated with being outwardly active and projected the brain's state onto this "sleep-wake" axis (Figure 5B). Figure 5B-right shows the distribution across participants of the center of various neurocognitive states along this axis, including stages of sleep as determined by an automated sleep score classifier (28). Neurocognitive states were consistently organized with the central attractor and wakeful rest separating actively waking behavior on one end and sleep stages N1, N2, N3, and REM on the other end. The organization of sleep stages in terms of distance from the central attractor matches the conventionally understood "depth" of sleep stages with deeper stages falling further away from the attractor (29). Stages of sleep broadly showed deactivation of anatomical networks across participants: REM sleep was associated with deactivation of the dorsal attention and salience network, while N3 and REM stages displayed several network deactivations that trended towards statistical significance (Figure S12).

Taken together, these results indicate that the default mode network serves an important role as an anchor in the brain's dynamics. Analogous to how heart rate increases during exercise but will on average have it trend back to resting heart rate via homeostatic forces, the brain's dynamics trend towards activation of the default mode network. This attractor along with neurocognitive states associated with wakeful rest and N1 sleep formed a central bridge between actively awake behavior and deeper stages of sleep.

Additionally, these findings suggest that long-standing results regarding the default mode network's role in resting-state fMRI not only generalizes to wakeful rest during natural behavior, but also forms a critical stabilizing anchor in neural dynamics (30).

Discussion

In this study, we investigated the expression, organization, and temporal dynamics of various neurocognitive states during natural human behavior. By projecting the data into a non-linear space where locations on this manifold represented the pattern of brain network activations, we found that neurocognitive states formed an hourglass-like structure. Behavioral states clustered in predictable locations along this structure, which allowed us to use network activation patterns to predict specific behaviors, such as whether they were talking to a friend or

watching a device. Awake outward behavior formed one end of the hourglass and sleep formed the other end. Times when participants were either awake but not outwardly active or in shallow sleep formed the central funnel.

The brain primarily switched its location along this structure by undergoing sharp bursts of dynamism where it would chaotically explore many areas before settling into a destination. These bursts tended to occur when someone's behavior was changing, such as when they went from looking at their smartphone to talking to their friends. The dynamics of brain network transitions suggest that when we switch our behavior in the real world, our brains do not undergo a stable, directed shift in neural activity but rather undergo an exploratory-like phase before stabilizing. Despite this chaos, the overall dynamics of the brain tended to be drawn towards a central homeostatic-like attractor located near the central funnel, wherein the brain tended to activate the default mode network and suppress sensory related ones.

Interspersed chaotic-like shifts are seen in other natural and computational systems, such as punctuated equilibrium in evolutionary biology or $1/f$ avalanches in cellular automata (31, 32). While these systems follow logical rulesets, they periodically generate complex and chaotic system-wide transformations such as phylogenetic explosions. Similarly, bursty transitions between stable states, known as "punctuated equilibrium," has been proposed as a hallmark of relatively efficient group decision-making (33). The "critical brain hypothesis" argues that emergent complexity in the brain, as in other large, multi-component systems, can occur by amplifying fluctuations on critical boundaries that follow characteristic power law dynamics with bursty characteristics, as we observed here (20, 34, 35). One common theme among these fields is the explore-exploit tradeoff: the concept that many systems incentivized to adapt to changing environments will alternate between exploration-heavy strategies that search for new solutions and exploitation-focused stratagems that fine-tune a single one (36). Our results indicate these frameworks may extend to continuous neural dynamics over longer timescales of hours to days during natural behavior.

Studying brain dynamics at this scale can enable the analysis of cognitive and physiological processes inaccessible on shorter timescales. Our attention, mood, and arousal fluctuate on the order of hours-to-days. Physiological changes such as hormones and gene expression do the same (37). Clinically, many neuropathological states evolve and fluctuate over this timescale. While these fluctuations can be difficult to assess using traditional experimental paradigms, there are ~600 thousand seconds of data in a single week of continuous recordings: 600 thousand examples of the brain's state in different behaviors, environments, and physiological conditions. Self-supervised deep neural networks, such as the one we used here, offer a rapidly developing method to detect patterns in sparsely labeled data, allowing us to link those patterns more accurately to behavior, physiology, and possibly pathological states. The Koopman operators we used here have seen increasing usage in control theory for their capability to identify underlying drivers of nonlinear system dynamics (38), a critical part of leveraging a system's natural dynamics during closed-loop control or modulation (9, 10, 39). To help facilitate the use of these methods for other applications, our analysis code can be found at <https://github.com/MNobodyWang/WeekLongBrain>.

Continuous human brain recordings over a week illustrate that brain networks transition between states via unpredictable and chaotic-like trajectories. These trajectories appear to explore many possible brain states before stabilizing into local states that predictably correspond to behavior, and activation of the "default mode network" during wakeful rest serves as a central attractor for the system. Taken together, these results suggest that the functional flexibility and

adaptiveness of our brains are an emergent property (40) of alternations between stable exploitation of specialized local brain states and wide exploration of the brain's possible configurations during state transitions, showcasing the utility of analyzing continuous neural recordings over long time periods during real-world behavior.

5 References and Notes

1. C. J. Honey, T. Thesen, T. H. Donner, L. J. Silbert, C. E. Carlson, O. Devinsky, W. K. Doyle, N. Rubin, D. J. Heeger, U. Hasson, Slow cortical dynamics and the accumulation of information over long timescales. *Neuron*. **76**, 423–434 (2012).
- 10 2. E. Simony, C. J. Honey, J. Chen, O. Lositsky, Y. Yeshurun, A. Wiesel, U. Hasson, Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature communications*. **7**, 1–13 (2016).
3. R. A. Poldrack, T. O. Laumann, O. Koyejo, B. Gregory, A. Hover, M.-Y. Chen, K. J. Gorgolewski, J. Luci, S. J. Joo, R. L. Boyd, others, Long-term neural and physiological phenotyping of a single human. *Nature communications*. **6**, 1–15 (2015).
- 15 4. E. M. Gordon, T. O. Laumann, A. W. Gilmore, D. J. Newbold, D. J. Greene, J. J. Berg, M. Ortega, C. Hoyt-Drazen, C. Gratton, H. Sun, others, Precision functional mapping of individual human brains. *Neuron*. **95**, 791–807 (2017).
5. L. H. Schulte, M. M. Menz, J. Haaker, A. May, The migraineur's brain networks: Continuous resting state fMRI over 30 days. *Cephalalgia*. **40**, 1614–1621 (2020).
- 20 6. D. Vidaurre, S. M. Smith, M. W. Woolrich, Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*. **114**, 12827–12832 (2017).
- 25 7. D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, S. T. Grafton, Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*. **108**, 7641–7646 (2011).
8. D. J. Newbold, T. O. Laumann, C. R. Hoyt, J. M. Hampton, D. F. Montez, R. V. Raut, M. Ortega, A. Mitra, A. N. Nielsen, D. B. Miller, Plasticity and spontaneous activity pulses in disused human brain circuits. *Neuron*. **107**, 580–589 (2020).
- 30 9. K. W. Scangos, G. S. Makhoul, L. P. Sugrue, E. F. Chang, A. D. Krystal, State-dependent responses to intracranial brain stimulation in a patient with depression. *Nature medicine*. **27**, 229–231 (2021).
10. S. A. Sheth, K. R. Bijanki, B. Metzger, A. Allawala, V. Pirtle, J. A. Adkinson, J. Myers, R. K. Mathura, D. Oswalt, E. Tsolaki, Deep brain stimulation for depression informed by intracranial recordings. *Biological Psychiatry*. **92**, 246–251 (2022).
- 35 11. M. Bijanzadeh, A. N. Khambhati, M. Desai, D. L. Wallace, A. Shafi, H. E. Dawes, V. E. Sturm, E. F. Chang, Decoding naturalistic affective behaviour from spectro-spatial features in multiday human iEEG. *Nature Human Behaviour*. **6**, 823–836 (2022).

12. A. Alasfour, P. Gabriel, X. Jiang, I. Shamie, L. Melloni, T. Thesen, P. Dugan, D. Friedman, W. Doyle, O. Devinsky, D. Gonda, S. Sattar, S. Wang, E. Halgren, V. Gilja, Spatiotemporal dynamics of human high gamma discriminate naturalistic behavioral states. *PLoS Computational Biology*. **18**, e1010401 (2022).
- 5 13. R. F. Betzel, J. Faskowitz, O. Sporns, Living on the edge: network neuroscience beyond nodes. *Trends in Cognitive Sciences* (2023).
14. L. Q. Uddin, B. Yeo, R. N. Spreng, Towards a universal taxonomy of macro-scale functional human brain networks. *Brain topography*. **32**, 926–942 (2019).
- 10 15. J. D. Murray, A. Bernacchia, D. J. Freedman, R. Romo, J. D. Wallis, X. Cai, C. Padoa-Schioppa, T. Pasternak, H. Seo, D. Lee, others, A hierarchy of intrinsic timescales across primate cortex. *Nature neuroscience*. **17**, 1661–1663 (2014).
16. C. H. Chang, S. A. Nastase, U. Hasson, Information flow across the cortical timescale hierarchy during narrative construction. *Proceedings of the National Academy of Sciences*. **119**, e2209307119 (2022).
- 15 17. D. R. Hardoon, S. Szedmak, J. Shawe-Taylor, Canonical correlation analysis: An overview with application to learning methods. *Neural computation*. **16**, 2639–2664 (2004).
18. A. Azarbarzin, M. Ostrowski, P. Hanly, M. Younes, Relationship between arousal intensity and heart rate response to arousal. *Sleep*. **37**, 645–653 (2014).
- 20 19. H. Zou, The adaptive lasso and its oracle properties. *Journal of the American statistical association*. **101**, 1418–1429 (2006).
20. P. Bak, *How nature works: the science of self-organized criticality* (Springer Science & Business Media, 1996).
- 25 21. D. Badre, A. D. Wagner, Computational and neurobiological mechanisms underlying cognitive flexibility. *Proceedings of the National Academy of Sciences*. **103**, 7186–7191 (2006).
22. D. R. Dajani, L. Q. Uddin, Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in neurosciences*. **38**, 571–578 (2015).
23. S. J. Schiff, K. Jerger, D. H. Duong, T. Chang, M. L. Spano, W. L. Ditto, Controlling chaos in the brain. *Nature*. **370**, 615–620 (1994).
- 30 24. K. Inoue, K. Nakajima, Y. Kuniyoshi, Designing spontaneous behavioral switching via chaotic itinerancy. *Science advances*. **6**, eabb3989 (2020).
25. H. Korn, P. Faure, Is there chaos in the brain? II. Experimental evidence and related models. *Comptes rendus biologiques*. **326**, 787–840 (2003).
- 35 26. M. L. Kringelbach, G. Deco, Brain states and transitions: insights from computational neuroscience. *Cell Reports*. **32** (2020).

27. B. O. Koopman, Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*. **17**, 315–318 (1931).
28. N. von Ellenrieder, L. Peter-Derex, J. Gotman, B. Frauscher, SleepSEEG: automatic sleep scoring using intracranial EEG recordings only. *Journal of Neural Engineering*. **19**, 026057 (2022).
29. M. A. Carskadon, W. C. Dement, Normal human sleep: an overview. *Principles and practice of sleep medicine*. **4**, 13–23 (2005).
30. M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, G. L. Shulman, A default mode of brain function. *Proceedings of the National Academy of Sciences*. **98**, 676–682 (2001).
31. S. J. Gould, N. Eldredge, Punctuated equilibria: an alternative to phyletic gradualism. *Models in paleobiology*. **1972**, 82–115 (1972).
32. Z. Olami, H. J. S. Feder, K. Christensen, Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. *Physical review letters*. **68**, 1244 (1992).
33. P. Bak, S. Boettcher, Self-organized criticality and punctuated equilibria. *Physica D: Nonlinear Phenomena*. **107**, 143–150 (1997).
34. L. Cocchi, L. L. Gollo, A. Zalesky, M. Breakspear, Criticality in the brain: A synthesis of neurobiology, models and cognition. *Progress in neurobiology*. **158**, 132–152 (2017).
35. W. L. Shew, D. Plenz, The functional benefits of criticality in the cortex. *The neuroscientist*. **19**, 88–100 (2013).
36. J. D. Cohen, S. M. McClure, A. J. Yu, Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **362**, 933–942 (2007).
37. O. Karin, M. Raz, A. Tendler, A. Bar, Y. Korem Kohanim, T. Milo, U. Alon, A new model for the HPA axis explains dysregulation of stress hormones on the timescale of weeks. *Molecular systems biology*. **16**, e9510 (2020).
38. A. Mauroy, Y. Susuki, I. Mezić, *Koopman operator in systems and control* (Springer, 2020).
39. S. Alagapan, K. S. Choi, S. Heisig, P. Riva-Posse, A. Crowell, V. Tiruvadi, M. Obatusin, A. Veerakumar, A. C. Waters, R. E. Gross, S. Quinn, L. Denison, M. O’Shaughnessy, M. Connor, G. Canal, J. Cha, R. Hershenberg, T. Nauvel, F. Isbaine, M. F. Afzal, M. Figee, B. H. Kopell, R. Butera, H. S. Mayberg, C. J. Rozell, Cingulate dynamics track depression recovery with deep brain stimulation. *Nature* (2023), doi:10.1038/s41586-023-06541-3.
40. M. Thiebaut de Schotten, S. J. Forkel, The emergent properties of the connected brain. *Science*. **378**, 505–510 (2022).

41. F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, R. M. Leahy, Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational intelligence and neuroscience*. **2011** (2011).
42. A. M. Dale, B. Fischl, M. I. Sereno, Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*. **9**, 179–194 (1999).
43. V. A. Traag, L. Waltman, N. J. Van Eck, From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports*. **9**, 1–12 (2019).
44. W. M. Rand, Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*. **66**, 846–850 (1971).
45. D. Pimentel-Alarcón, R. Nowak, "Random consensus robust PCA" in *Artificial Intelligence and Statistics* (PMLR, 2017), pp. 344–352.
46. P. van Gent, H. Farah, N. Nes, B. van Arem, "Heart rate analysis for human factors: Development and validation of an open source toolkit for noisy naturalistic heart rate data" in *Proceedings of the 6th HUMANIST Conference* (2018), pp. 173–178.
47. C. Truong, L. Oudre, N. Vayatis, Selective review of offline change point detection methods. *Signal Processing*. **167**, 107299 (2020).
48. G. A. Gottwald, I. Melbourne, The 0-1 test for chaos: A review. *Chaos detection and predictability*, 221–247 (2016).
49. A. Clauset, C. R. Shalizi, M. E. Newman, Power-law distributions in empirical data. *SIAM review*. **51**, 661–703 (2009).
50. E. Yeung, S. Kundu, N. Hodas, "Learning deep neural network representations for Koopman operators of nonlinear dynamical systems" in (IEEE, 2019), pp. 4832–4839.

Acknowledgments: We would like to thank the patients for their participation in this research study. We also thank the physicians and staff in the University of Pittsburgh Comprehensive Epilepsy Center at the University of Pittsburgh Medical Center, particularly Cheryl Plummer and Anto Bagic, without whom none of these data exist. We thank T. Verstynen, R. Kass, L. Holt, T. Abel, A. Alreja, and D. Geng for comments on the manuscript.

Funding: This work is supported by grants from the National Institutes of Health (R01MH132225, R01MH107797, P50MH109429) and NSF (1734907). M.B.W. is supported by the Hertz Foundation and NIH T32GM008208.

National Institutes of Health grant R01MH132225, R01MH107797 (ASG)

National Institutes of Health grant P50MH109429 (ASG)

National Institutes of Health grant T32GM008208 (MBW)

National Science Foundation grant 1734907 (ASG)

Hertz Foundation (MBW)

Author contributions:

Conceptualization: ASG, MBW
Methodology: MBW, MG, JFC, ASG
Investigation: MBW, ASG, RMR
Visualization: MBW, ASG
Funding acquisition: ASG, MBW
Project administration: ASG
Writing: MBW, ASG

Reviewing and Editing: MBW, MG, JFC, RMR, ASG

Competing interests: Authors declare they have no competing interests.

Data and materials availability: Code available on <https://github.com/MNobodyWang/WeekLongBrain>. Data available on reasonable request.

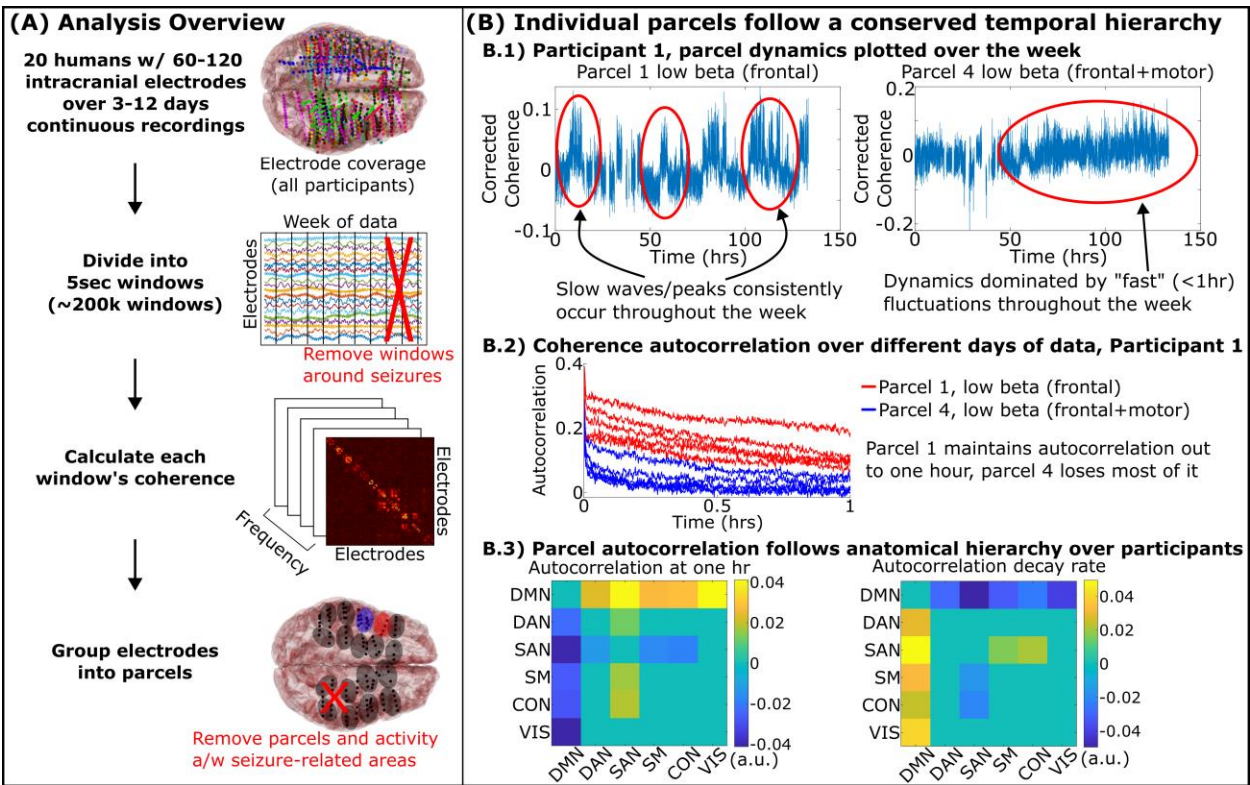
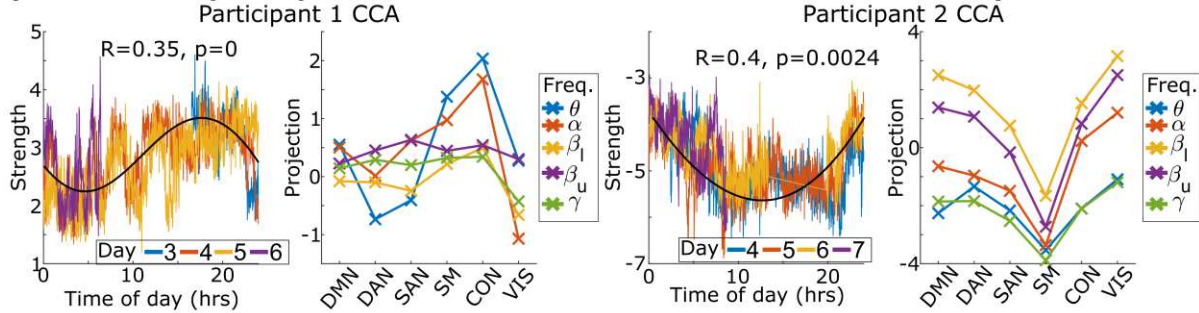


Fig. 1. Parcels of the brain followed stable rhythms conserved throughout the week that followed an anatomical hierarchy. A) 3-12 days of continuous recordings from twenty participants were split into five-second-long windows, removing windows around seizure activity and artifact removal. We calculated the coherence between all pairs of electrodes and grouped electrodes with high coherence into anatomically compact parcels. B.1) Coherence within two parcels from a representative participant. B.2) Parcels display unique, stable timescales reflected by their autocorrelation stability over different days of data (all participants shown in Supplementary Figure S3). B.3) Timescale of rhythms between parcels belonging to

one fMRI resting network versus another. Cell values indicate the difference in autocorrelation parameter across participants (y axis versus x axis) with positive cells indicating the network indicated by its row has a larger parameter than the one indicated by its column. Non-zero cells indicate statistically significant differences post multiple comparisons by mixed effects model. Methods described in Supplementary Section M6.

A) 11 out of 20 participants had networks that correlated to circadian rhythm



B) 6 out of 7 participants had networks that predicted heart rate

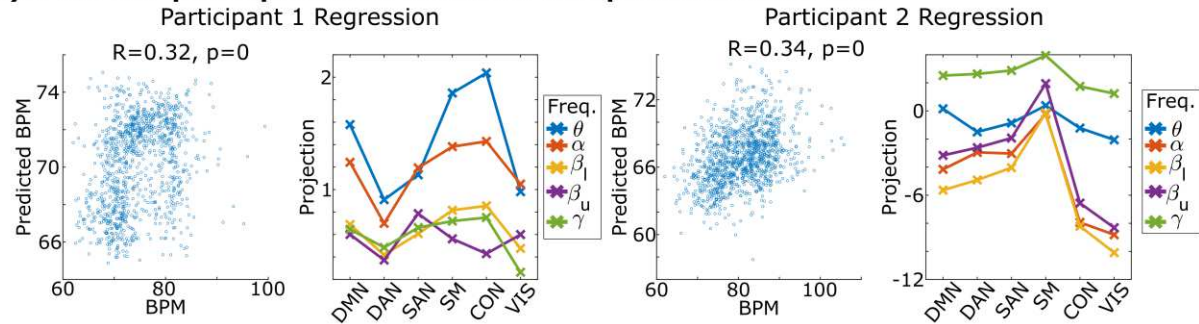


Fig. 2. Brain networks predicted physiological markers. A) We linked networks (groups of parcels found using principal components analysis) to circadian rhythm by training canonical correlation analysis on one half of the week and then testing on the other. The network mixture activations during testing are shown on the left plotted against time with the black line indicating a theoretical circadian rhythm. Skips in data are removals due to seizures or disconnected hardware. The identified mixture's anatomical and frequency coverage are shown projected onto the canonical fMRI networks. B) Networks were linked to heart rate by training linear regressors on one half of the week and testing on the remaining half. Test predictions are plotted against heart rate along with their anatomical and frequency coverage. All participants for these analyses are shown in Supplementary Figures S5 and S6. Methods described in Supplementary Section M9.

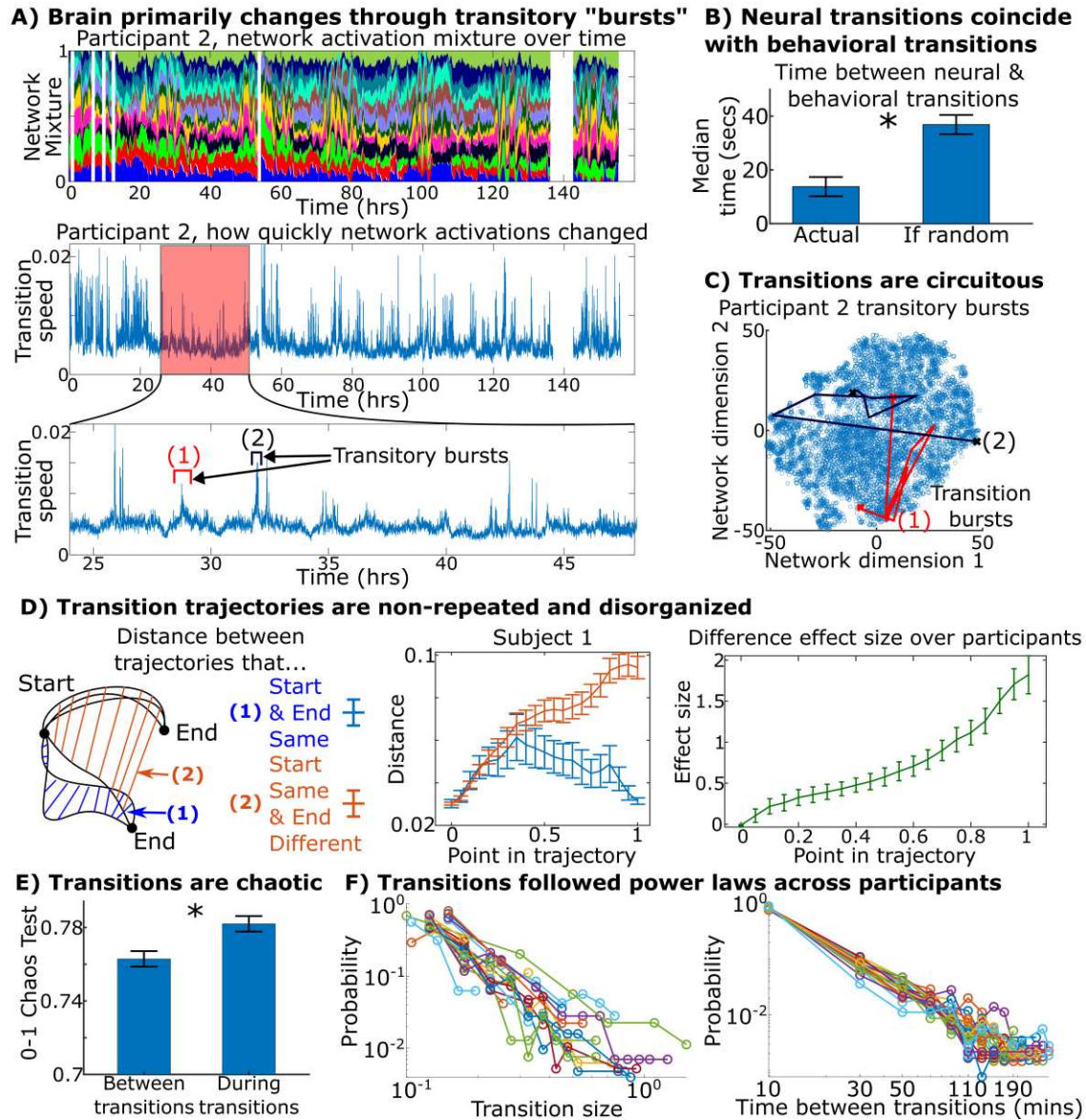


Fig. 3. Neural dynamics undergo chaotic-like transitions when natural behavior shifts. (A-top) Network activations plotted over the week for one participant where each color represents the activity of a different network (sum of activity normalized to one for visualization purposes). (A-mid/bottom) How quickly the brain changed network activations every five seconds. The brain reorganized itself using “transition bursts” of high speed (Supplementary Figure S7 quantitatively demonstrates that transitions are “bursty” across all participants). (B) Average time across participants between neural and behavioral transitions compared to the expected time if no relation between the two. Neural and behavioral transitions tended to occur with each other ($p=1e-4$, paired t-test). (C) Two transitions visualized on a t-distributed stochastic neighbor embedding of the brain’s weeklong course, showing that transitions did not move directly between states but rather explored many interim states (quantified in Supplementary Figure S13 over participants). (D) We took transition bursts with the same starting and ending states and asked how similar they were as transitions progressed from start to end (1, blue). We compared this to transitions with the same starting but different ending ones (2, red). The Cohen’s d effect size on the difference between these two distances is shown on right. The first half of the transitions indicated little about the eventual destination, indicating that transitions in the brain

did not take consistent paths from start to end. (E) 0-1 chaos test shows that the chaoticity of brain dynamics rises during transitions across participants ($p=1e-3$, paired t-test). (F) Distribution of transition size and the time between them for all participants in log-log form. Both distributions formed power laws (linear on log-log axes) across participants by Kolmogorov-Smirnov and likelihood tests. Details in Supplementary Section M10.

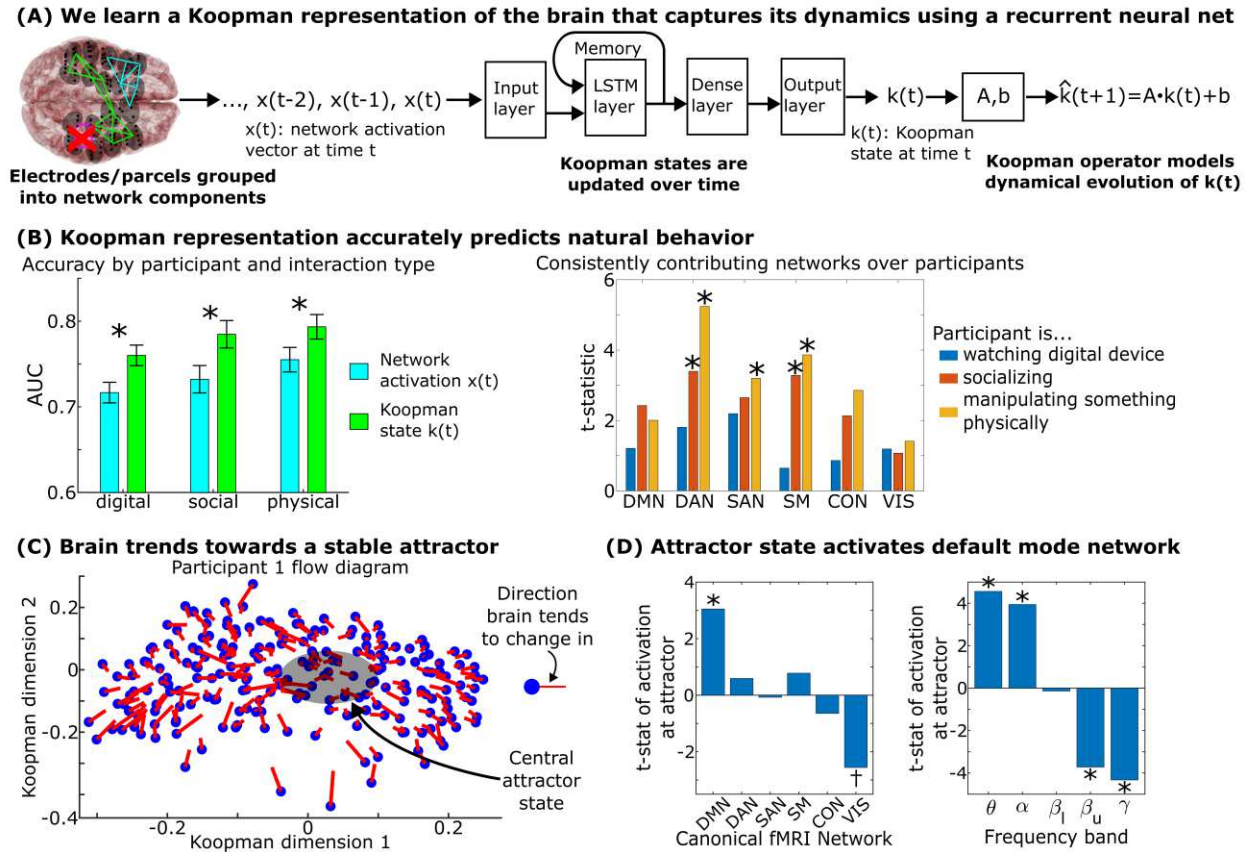
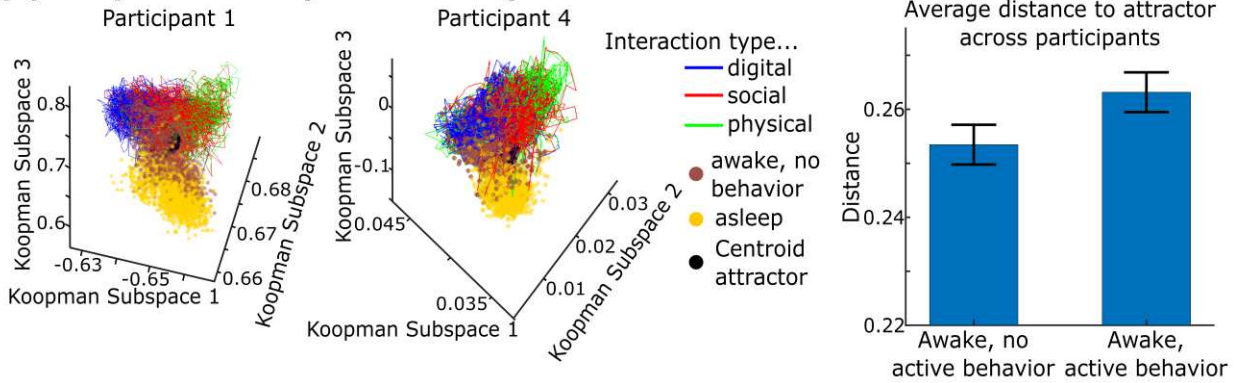


Fig. 4. Neural dynamics are driven by a central homeostatic-like attractor at the default mode. (A) We learned a Koopman representation of the brain's dynamical state by using a recurrent neural network to project the original network activations into a higher-dimensional "Koopman space" where the trajectories of the brain in this space could be captured by linear operators. (B) Trajectories in Koopman space more accurately predicted natural behavior than the original network activations ($p=0.012$ by paired t-test). Error bars on the left indicate 95% confidence intervals across participants. Anatomical regions consistently activated during each behavior across participants are shown on the right. (C) Flow diagram of how the brain's state tends to change over time, showing their overall tendency to drift towards a central attractor state which is quantified over all participants in Supplementary Figure S11. (D) At this central attractor state, the brain consistently activates the default mode network at low frequencies across participants ($p<0.05$ post multiple comparisons correction). The dagger marks a t-statistic that is significant independently ($p=0.02$) but not significant post multiple comparisons correction. Methods described in Supplementary Section M11-13.

(A) A day of neural trajectories during natural behavior



(B) Attractor lays at the middle point between sleep-wake axis

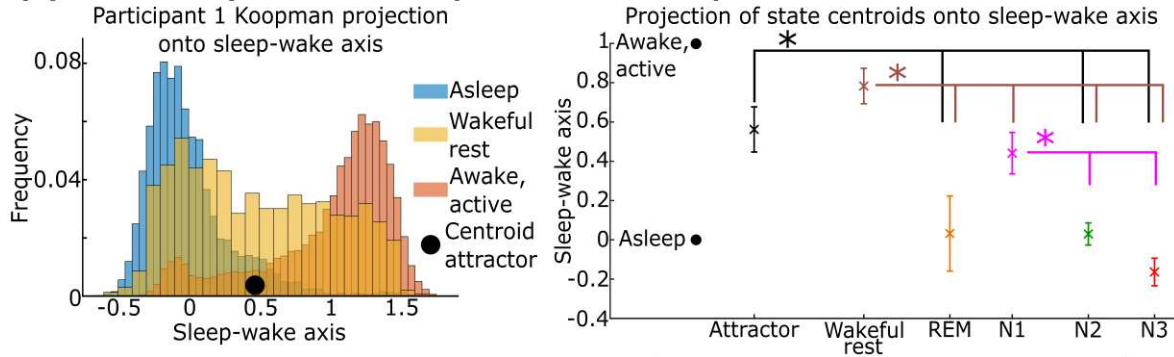


Fig. 5. Wakeful rest orbits the central attractor while active behavior departs it. (A-left)

This plot shows what brain networks are doing over the course of a day in two participants.

Specifically, we plotted a full day of brain network trajectories in the Koopman space,

highlighting both the central attractor state along with how different wake states and sleep are

positioned relative to this attractor. (A-right) We found that times where the participants were

doing one of the three active behaviors tended to depart further away from the central dynamical

attractor state relative to times when the patient was awake but not doing any of the three

behaviors ($p=0.03$ by paired t-test). (B-left) We projected neural states from one participant onto

the axis separating the centroid of their sleeping and waking states, which we denote the sleep-

wake axis. This is equivalent to projecting the hourglasses shown in Figure 5A-left onto a

vertical line nearly parallel to the z-axis that runs between the center of asleep and waking states.

(B-right) We calculated the centroid of each participant's neurocognitive states and projected

them along their sleep-wake axis. The center of sleeping states was normalized to zero, and the

center of actively awake states was normalized to one. We then plotted the centroids' distribution

across participants (error bars indicate 95% confidence intervals). Overhead asterisked bars

indicate that the state indicated by the color was significantly different from the marked states by

multiple comparisons corrected paired t-tests ($p<0.05$). We found that wakeful rest and the

associated centroid attractor occupied a middle position in the sleep-wake axis with different

sleep stages naturally organizing according to their conventionally understood depth. Anatomical

regions consistently associated with each sleep stage are illustrated in Figure S12. Methods

described in Supplementary Section M14.

Supplementary Materials for

A week in the life of the human brain: stable states punctuated by chaotic transitions

5 Maxwell B. Wang, Max G'Sell, James F. Castellano, R. Mark Richardson, and Avniel Singh
Ghuman

Corresponding author: mbwang@alumni.cmu.edu; ghumana@upmc.edu

The PDF file includes:

10

Materials and Methods
Figs. S1 to S12

Materials and Methods

Methods

M1. Participants

Twenty participants (nine males, eleven females; mean age 40 years with a standard deviation of 12 years) had intracranial surface or depth electrodes implanted for the treatment of pharmacologically intractable epilepsy (Figure 1A). All participants gave informed consent to participate under research protocols approved by the University of Pittsburgh Institutional Review Board. Depth electrodes were produced by Ad-Tech Medical and PMT and were 0.86 and 0.8 mm in diameter, respectively. Grid electrodes were produced by PMT and were 4 mm in diameter. Sixteen participants had depth electrodes only, three had grids only, and one had a combination of both. The main results of this paper, when studied on an individual participant by participant basis, did not differ significantly between those with depth electrodes and those with grids.

M2. Analysis overview

In summary, we collected (M3) and preprocessed chronic intracranial recordings continuously over multiple days, which we divided into five second non-overlapping windows, and calculated the functional partial connectome of each window via all-to-all electrode coherence (M4). The electrodes were grouped into tightly connected parcels (M5) and the timescales (autocorrelation) of the parcels determined (M6). We then grouped parcels and frequencies into functional network components using Principal Components Analysis (M7) and studied how the overall mixture of all functional networks would change over time (M9/10). Finally, we learned a dynamical systems model of these networks using a deep recurrent neural network to study the attractor and repulsor dynamics governing brain network dynamics and the relationship between brain networks and behavior (M11-14).

Artifacts were removed at multiple points in the analysis. Specifically, a comb filter was applied to remove line noise (M3). An hour before, during, and after all seizures were removed to eliminate ictal and peri-ictal activity (M4). Spatial regression was used to remove motion, respiratory, and cardiac artifacts (M4). ICA was used to remove large spike artifacts that sometimes occur due to disturbing the cables or connections (M4). Epileptogenic areas and activity that correlated with the activity in these regions was removed to eliminate interictal activity or other pathological activity (M4/5/8).

M3. Intracranial EEG data collection

Electrodes were localized via postoperative MRI or CT scans coregistered to the preoperative MRI using Brainstorm (41). Surface/grid electrodes were projected to the nearest point on the preoperative cortical surface automatically parcellated via Freesurfer to correct for brainshift (42). Electrode coordinates were then coregistered via surface-based transformations to the fsaverage template using Freesurfer cortical reconstructions. Intracranial electroencephalography (iEEG) data was collected using the Natus system at 1kHz. Notch filters at 60 and 120 Hz were applied with a subsequent bandpass filter from 0.2 to 115Hz.

M4. Data preprocessing and artifact removal

The spatial autocorrelation between an electrode and all electrodes within 2cm of it was then measured and regressed out to remove both local and global artifacts, including artifacts due to motion and current spread due to volume conduction.

Neural activity an hour before and an hour after all seizures, whether electrographic or clinical, was removed before calculating coherence. A board-certified neurologist identified the seizure network in all but two participants, with those two participants having no recorded electrographic or clinical seizures during their stay in the hospital.

The data was then separated into five second non-overlapping windows and coherence was computed over each window between all pairs of electrodes over five frequency bands: theta (4-8Hz), alpha (8-12Hz), low beta (14-20Hz), high beta (20-30Hz), and gamma (30-70Hz). In summary, this generated five connectome structures every five seconds. This was performed using the scipy coherence function under default settings as of version 1.9.3.

Independent component analysis was then applied, and components were visually inspected for any artifacts which were then removed. The main criteria for removal were independent components that possessed time course activations that were clearly non-neurological (such as step-functions or near Dirac deltas).

M5. Parcellation

For each participant, we parcellated their electrodes into groups of tightly coherent electrodes. We utilized the Leiden algorithm to identify a single regional atlas that optimized graph modularity over the entire week-long period across all five frequency bands (43). Modularity (Equation 1) was calculated separately over each network from every five-second window with the Leiden algorithm optimizing the average modularity across all windows and frequencies. This generated on average 10-15 parcels for each participant.

$$\text{modularity} = \sum_{b \in \{\theta, \alpha, \beta_l, \beta_u, \gamma\}} \sum_t \sum_{i,j} \left[A_{i,j}^{b,t} - \frac{k_i^{b,t} k_j^{b,t}}{2m^{b,t}} \right] \delta(c_i, c_j)$$

Equation 1: Modularity metric that the Leiden algorithm optimizes. $A_{i,j}^{b,t}$ refers to the weighted connectivity (coherence) between electrodes i and j at time window t and frequency band b . $k_i^{b,t}$ is the degree of electrode i and $m^{b,t}$ is the sum total of all connections at that time and frequency. $\delta(c_i, c_j)$ is an identity function which denotes whether electrodes i and j are in the same parcel. The Leiden algorithm finds the parcellation assignment of each electrode that optimizes modularity over all time windows and frequency bands.

To assess the stability of which electrodes would be grouped into which parcels, we separated the data into six-hour non-overlapping segments (between 18-80 segments per participant) and found the optimal community structure for each segment. We quantified the similarity between each segment's parcel definitions using the Rand Index (44) (percentage of electrode pairs that were parcellated equivalently under the two parcel definitions) which almost universally returned values greater than 0.9 as illustrated in Supplementary Figure S2 indicating that the overall parcellation was well-preserved over time. Across participants, this average value had a mean of 0.96 and a standard deviation of 0.02, indicating parcellations remained highly consistent over time. This consistency motivated our decision to use the same parcel structure over the entire work for interpretability.

To ensure that that any potential pathological activity was removed from our analysis, we removed elements of the activity of non-seizure related parcels that appeared to correlate to the activity of parcels within the seizure onset zone. More specifically, using linear regression, we attempted to predict the coherence within non-seizure related parcels using the coherence of seizure-related parcels and then removed this predicted coherence.

M6. Autocorrelation stability (Figure 1)

We tested whether the autocorrelation of each parcel's coherence would show relatively consistent patterns of “fast” or “slow” rhythms throughout the week (Figure 1). We split the entire ~week-long time course for each participant into six-hour non-overlapping segments. After removing parcels associated with the seizure network, we then took the average coherence between electrodes within a single parcel for a single frequency band and then calculated its autocorrelation up to one hour. We fit this autocorrelation curve to a power law ($\text{autocorrelation}(t) = AC_1 \times \text{time}^{-AC_2}$) to generate two timescale parameters: AC_1 (autocorrelation strength) and AC_2 (autocorrelation steepness) which described the autocorrelation of a single parcel at a single frequency at a single time segment. For a given frequency band, we took the timescale parameters across all parcels and time segments and grouped the parameters by which parcel they were measured in. We used Kruskal-Wallis ANOVA tests to show that in almost all participants and frequency bands, there were statistically significant differences between the group means, mostly with high effect sizes ($\eta > 0.12$, Supplementary Figure S3).

We tested whether parcels from different anatomical regions tended to have reliable differences in their autocorrelation across participants using linear mixed effect models. We assigned each parcel to one of the six canonical fMRI networks defined in (14) (“default mode”, “dorsal attention”, “salience”, “somatomotor”, “control”, and “visual”) as based on its largest overlap. For each parcel, we calculated the autocorrelation of its average intra-parcel coherence for a given frequency over the entire week out to one hour and calculated AC_1 and AC_2 as described above. We then averaged both parameters across all frequency bands.

We then chose a single pair of fMRI networks (such as “salience” vs “visual”) and selected all the parcels across our participants that fell into one of those two anatomical groups. We used MATLAB's fitlme (linear mixed effect model) to model each parcel's autocorrelation parameters with the anatomical group as a fixed-effect and the participant as a random effect, allowing us to determine whether one anatomical group had a reliably higher autocorrelation parameter than the other across participants. We repeated this for all possible pairs of fMRI networks and used Bonferroni multiple comparisons correction to identify pairs with significant differences (Figure 1B.3).

M7. Robust principal components analysis

Since many parcels tended to be highly colinear, we used a modified PCA protocol to reduce dimensionality. We grouped parcels and frequencies that tended to covary together using random sample consensus PCA (RANSAC-PCA) on the parcel coherences. By taking the average intra-parcel coherence during each time window and frequency, we formed a (number of parcels x 5 frequency bands) by (number of time windows) 2D matrix which we then reduced to a (number of components) by (number of time windows) matrix using the modified PCA protocol. This identifies parcels and frequencies that tend to strongly covary together that we could easily interpret as a single network component feature that captures cross-frequency

relationships while also reducing the dimensionality of the original dataset to simplify further analyses.

The modified PCA protocol we used applies random sample consensus to avoid PCA's susceptibility to noisy outliers by taking multiple small subsamples of the data and selecting one with the fewest number of outliers to train the model (45). We generated 1000 subsamples where in each subsample, we selected six 30-minute segments of data from each day. This ensures that the PCA is robust to rare outliers and that the PCA produces principal components that are stable across days within a participant. Outliers were defined by calculating the Mahalanobis distance between each time window's feature vector and each subsample's distribution. In each participant, we found that these distances would take on clear bifurcations between relatively small distances and short "spikes" of extremely high distances (more than three standard deviations) away from the mean that typically lasted for a few minutes. We manually drew a cutoff for each participant that was approximately half the average Mahalanobis distance of these spikes. For each subsample, we calculated the number of outliers within the subsample, and calculated PCA over the subsample with the fewest outliers. We utilized enough PCs to capture 90% of the variance in the dataset, generally resulting in 12-24 networks/PCs per participant.

The network component activation of a principal component was defined as the projection of the parcel coherences onto the principal component weights.

M8. Seizure network removal

When analyzing parcel dynamics (Figure 1), we excluded all parcels with electrodes part of the seizure onset zone and early propagation as defined by a board-certified neurologist. For network component dynamics (Figure 2 onwards), we first re-added these seizure-related areas before grouping parcels and frequencies into network components through robust PCA. We then removed any network components that were associated with the seizure network before analyzing their dynamics. More specifically, we calculated the dot product similarity between the absolute value of a principal component vector (normalized to a magnitude of one) and a binary vector that marked all electrodes that were part of the seizure network (also normalized to one). The similarity between these two vectors indicated how anatomically similar the driving factors of a principal component and the seizure network were to each other. A null distribution for this similarity was formed by randomly permuting the principal component vectors, and all principal component vectors that showed statistically significant similarity to the seizure network ($p < 0.05$) were removed from all further analyses.

M9. Network activation is tied to circadian rhythm and heart rate (Figure 2)

We tested whether we could identify combinations of networks that were associated with neurophysiologically relevant markers. More specifically, we looked at circadian rhythm and heart rate.

Canonical correlation analysis (CCA) was used to identify a mixture of network components that matched a circadian sinusoid with a period of 24 hours. The circadian sinusoid was defined as $a1 * \cos(t/24\text{hrs}) + a2 * \sin(t/24\text{hrs})$ where $a1$ and $a2$ are constants learned via CCA. CCA simultaneously tried to find a linear combination/weighting of network component activations to fit to this sinusoid.

The model was trained over the first half of the week and then tested on the second half through Pearson correlation (out of sample validation of correlation). The Pearson R of the fit on the test dataset was calculated and then compared to a null distribution of R that was formed via permutation tests that temporally shifted each day's network component activity forward or backwards by a uniform random number ranging from 0-24 hours. This preserves the autocorrelation of the neural signals while eliminating any consistent circadian-like pattern across days.

Heart rate was assessed using collected EKG signals that were processed using heartpy (46). The instantaneous heart rate for any window was the average heart rate for a 30-second period centered on the window. L1-regularized regression was trained on the first half of the week to identify a mixture of networks that predicted heart rate using sklearn's implementation (out of sample validation of regression). Hyperparameters were optimized on the training set using ten-fold cross-validation. The quality of the fit was assessed on the remaining half of the week via Pearson correlation with a null distribution formed using the same permutation tests used for circadian rhythm to preserve both the autocorrelational properties of the heart rate and neural signals.

M10. Neural dynamics undergo chaotic-like transitions that are associated with shifts in natural behavior (Figure 3)

We examined how the overall brain state (the status of all recorded networks in the brain) would change over time by dividing the week into “transitions”, periods when the brain was rapidly reconfiguring itself; and “states”, periods of time where the brain's functional connectome appeared to be relatively stable.

In Figure 3A and Supplementary Figure S7, we provide evidence that the brain falls into states and transitions by examining the “speed” of the brain. Speed was defined as how much the brain's state changed between one five-second window and the next. More specifically, we took the vector of all network activations of each window (the parcel coherences projected into the network PCA space) and calculated the Euclidean distance between the network activation vector of one window and the next (the temporal derivative in the network PCA space).

We calculated the distribution of the time between windows that fell into the top 1% of velocity across all participants and compared that distribution to Poisson distributions with $\lambda=0.01$. The Poisson distribution captures what the expected time between high-speed windows would be in a memoryless process (non-autocorrelated speed). The results showed that windows with high speeds tended to cluster next to each other temporally (supplemental Figure S7). This finding supports the notion that brain network dynamics can be separated into times when the brain is relatively static (states) punctuated by bursts of time when the brain is quickly changing (transitions). Furthermore, we tested this between windows falling within the top 10% of velocity ($\lambda=0.1$) and found the same result. On an individual participant level, we also used Kolmogorov-Smirnov tests to assess whether the time between high transition speed windows followed a Poisson distribution, rejecting the null hypothesis in all 20 participants ($p<0.05$).

In Figure 3D-F, we evaluated various properties of these transitory bursts. We identified neural transition bursts using binary segmentation on the time series of network activation vectors (change point detection), implemented via the ruptures package in Python using default settings (47). This algorithm starts by identifying the optimal place to position a single change

point such that the network activations before the change point possess a maximally different distribution from the network activations after it, effectively dividing the week into two pieces. Binary segmentation then recursively subdivides the resulting pieces of time according to the same procedure, until no such change points can be found. Neural transition trajectories were then defined as the time around a change point possessing above average transition speed.

In Figure 3B, we calculated the median time between neural and behavioral changepoints. Using nine participants with available high-quality video recordings throughout the week, we randomly selected two days to annotate for one of three natural behaviors: interfacing with a digital screen, socially interacting with another living creature (either human or in one participant a canine companion), or physically interacting with an object. For this analysis, we excluded times when participants were asleep. In practice, if the participant was not doing one of these three behaviors, they were awake but not outwardly active (wakeful rest/presumably internally thoughtful). Behavioral changepoints were defined as anytime one of these three behaviors started or ended, including times when participants switched from one to another. We calculated the median time between neural and behavioral changepoints and found a null distribution for this metric using permutation testing.

Specifically, we temporally shifted the behavioral changepoints forward or backwards by a uniform random number ranging from 0-24 hours, recalculating the median time difference between behavior and neural changepoints, and used 10000 trials of this to form a null distribution of the expected time difference between behavioral and neural changepoints if there was no temporal relation between the two. We tested whether the real time difference between behavior changes and state changes was consistently smaller than the expected time difference using a paired t-test across participants.

In Figure 3C, we examined the distance between the paths traversed by different transition trajectories. We calculated the distance between the start points of all trajectories in a participant and the distance between their end points. Two trajectories were considered to have similar starting or ending point if the distance between the points fell into the bottom 10% of trajectory pairs. We grouped trajectories into three groups: 1. trajectories with similar starting and ending point, 2. trajectories with similar starting points only, and 3. trajectories with similar ending points only. We calculated the average distance between trajectories that fell into each group as a function of how much of the trajectory had been completed. Specifically, we used linear interpolation to determine what was the brain state 5%, 10%, 15%, 20%, ..., 95% of the way into each trajectory. We calculated the distance between brain states of the same percentage in each of the three groups. Figure 3D shows the distribution of these distances for a single participant and the effect size of the difference between these distances across all participants. To determine the effect size, for each participant, we calculated the Cohen's d between the distances between trajectories that start and end similarly to the distances between trajectories that start similarly but end differently. This measures the number of standard deviations that separate the distributions in the trajectory categories at different points along the trajectory. We then calculated the standard error of these Cohen's d across all 20 subjects.

In Figure 3E, we studied whether these transitions influenced the chaoticity of the brain dynamics. Chaoticity was defined using the 0-1 chaos test using the protocol described in (48) and was calculated over non-overlapping ten minute segments. In summary, we calculated the chaoticity of each network component independently over each time segment. The chaoticity of

the overall neural dynamics for a given time segment was defined as the median chaoticity of all network components. Segments with a transition were compared to segments without one.

$$p(n+1) = p(n) + \phi(n) \cos cn \quad (2.1)$$

$$q(n+1) = q(n) + \phi(n) \sin cn \quad (2.2)$$

$$M(n) = \lim_{n \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N ([p(j+n) - p(j)]^2 + [q(j+n) - q(j)]^2) \quad (2.3)$$

$$K_c = \lim_{n \rightarrow \infty} \frac{\log M(n)}{\log n} \quad (2.4)$$

Equation 2: Chaos 0-1 test protocol. Define $\phi(n)$ as the network component activation of interest at time window n for a given ten-minute segment. This is used to “drive” the dynamical system described in Equations 2.1 and 2.2 where c is a randomly chosen “resonance” parameter between 0 and π that remains constant during a single “iteration” of this process. $M(n)$ is evaluated up to an n of approximately $N/10$ where N is the number of time windows in the ten-minute segment. K_c is estimated by fitting a straight line between the numerator and denominator of Equation 2.4 and represents the chaoticity of a single iteration. c is redrawn 1000 times and the median K_c is defined as the chaoticity of the network component over the ten-minute segment.

In Figure 3F, we analyzed the distribution of the transition size which we defined as the net displacement of a transition and the time-between transitions. We fit power law exponents to these distributions using MATLAB’s `nlinfit` function with power laws defined as $a1 * \text{frequency}^{(-a2)}$ where $a1$ and $a2$ are learned.

We tested whether these distributions came from power law distributions using two methods from (49). First, we used Kolmogorov-Smirnov (KS) tests to test whether we failed to reject the null hypothesis that the distributions plausibly came from power laws. We fit power law distributions to each participant’s transition size and time between transitions distributions separately and calculated the KS distance between the experimental distributions and their theoretical power law distributions. We formed a null distribution on these distances by drawing 10000 random samples from the theoretical power law distribution, fitting a power law distribution to those samples, and then calculating the KS distance between the sampled distribution and the fitted one. If these distances were consistently lower than the distance between the real distribution and its estimated power law one, then we reject the null and conclude that the distribution did not come from a power law. We found that 17/20 participants had transition size distributions that plausibly came from power law distributions ($p > 0.05$), and 20/20 participants had time-between transition distributions that plausibly came from power law distributions.

We then used likelihood comparison tests to see whether the transition size and time-between distributions were more likely to have come from power law, exponential, or log-normal distributions. We calculated the log-likelihood that each participant’s distributions came from each of the three categories. We used a Wilcoxon signed-rank test to assess whether the log-likelihood of power law distributions were higher than exponential and log-normal distributions across participants. Power law distributions were more likely than exponential

($p=0.007$ for transition size and $p=4.8e-5$ for time-between) and more likely than log-normal ($p=0.03$ for transition size and $p=4.8e-5$ for time-between).

M11. Neural dynamics are driven by a central homeostatic-like attractor at the default mode.

To determine whether these brain network dynamics possessed consistent anatomical trends, we used dynamical systems eigendecomposition.

We used a deep recurrent neural network to learn a high-dimensional representation of the network activations where the brain's dynamics could be captured using linear methods. The full model is shown in Figure S9. Our model is an adaptation of the method presented in (50) which uses deep neural networks to learn a nonlinear transformation between the original observables of a system onto their corresponding points on a new manifold where the evolution of the system can be captured by linear dynamical laws.

To describe a single forward-pass through the model, we start with x_t , the network activation vector at time t (the intraparcels coherences projected onto the robust principal components) for a single participant. This vector was first fed into an encoder, a long short-term memory (LSTM) unit implemented in tensorflow using default settings.

The output of the LSTM was the Koopman representation k_t , a set of variables that summarized the current dynamical state of the brain up to time t . k_t was chosen to have a dimensionality ten times as large as x_t to allow it to serve as a nonlinear kernel. k_t was then passed to two models during training. The first was a linear autoregression model with learnable parameters A , b that attempted to predict the next time step's Koopman state as $\hat{k}_{t+1} = Ak_t + b$. The autoregressive error of this is the squared loss between the predicted and actual Koopman state. In other words, how well does the information encoded in k_t predict its own temporal evolution using linear methods?

We trained all models (f , d , A , b) simultaneously according to two loss functions. One of these loss functions asked whether the predicted Koopman state representation at time $t+1$ was close to the actual Koopman state representation at that time. A single training step is described in Figure S10.

We implemented these models using Python's tensorflow Adam optimizer under default settings. All networks used activation functions and L-1 regularization.

M12. Behavioral classification (Figure 4B)

To test whether the Koopman state representation captured neurocognitively interesting information and whether behavior organized consistently along it, we used it to predict the participant's behavior. Participant behavior was manually annotated for three activities: watching a digital screen, socializing with someone else, or physically manipulating an object. These three behaviors were not mutually exclusive. We trained L1-regularized logistic classifiers using Python's sklearn toolbox to predict behavior by training on one day and testing on another (binary classifiers, were they performing the behavior or not). Hyper parameterization was optimized on the training set using ten-fold cross-validation. The area-under-curve of the

receiver-operator-characteristic of each network's ability to classify the desired behavior was calculated.

To determine what brain networks were consistently associated with each behavioral state across participants, for each participant, we selected time windows where the participant was doing a behavior of interest. Using the decoding model, we asked what brain networks were associated with the Koopman state representation of those time windows, forming a (networks x 1) vector for each time window. We averaged these vectors across all time windows associated with the behavior and projected them onto the six canonical fMRI networks. Using the temporal perturbation testing described in our behavioral changepoint testing (temporally shifting the data forwards or backwards by random amounts relative to the behavioral labels), we formed a null distribution for the activation of these six canonical fMRI networks. We then calculated the Cohen's d effect size of the actual average network activation during a task compared to the null distribution with positive effect sizes indicating higher-than-random network activation. Across all participants, we then used one-sample t-tests to see if there was a consistent trend in these effect sizes across participants and used Benjamini-Hochberg for multiple comparisons correction (with the number of tests being 18 for three behavioral tasks over six networks).

M13. Attractor state analysis (Figure 4C)

We calculated the eigendecomposition of the Koopman operator (A from Figure S9). In all subjects, A was full rank and non-ill conditioned, leading to one found critical point (also known as a fixed or equilibrium point): $\text{inv}(I-A)*b$ where I is the identity matrix. We found that in all twenty participants, the real component of all eigenvalues had magnitude greater than zero and less than one, indicating this critical point was an attractor (brain states near the attractor tend to be pulled towards it) as shown in Supplementary Figure S11. While dynamical systems may have multiple critical points, this single attractor found by the Koopman operator is conventionally thought to represent the “global” behavior of the system (e.g. a system with both an attractor and repeller will show eigenfunctions associated with the attractor if the system gravitates towards the attractor as time approaches infinity (14)). One caveat is that interpreting systems with more than one critical point using Koopman operators is an active area of investigation (15).

We used the decoding model to ask what original network activation was associated with this attractor state. We projected the resulting network activations onto the six canonical fMRI networks as defined in (16) and averaged over frequency, giving us a six-by-one vector for each participant. We then subtracted each participant's fMRI network activation vector by its mean to ask “which networks were activated or deactivated” relative to the rest of the brain. We used t-tests on each network activation across participants to see if any networks were consistently activated or deactivated. We used Bonferroni for multiple comparisons correction. Participants that did not have electrodes in all six canonical networks were removed from this analysis.

We repeated this process except by averaging over networks to ask if any frequencies were activated or inactivated.

M14. Neurocognitive states form an hourglass-like shape where the default mode attractor separates waking and sleep (Figure 5)

For visualization purposes, in Figure 5A-left, we chose two participants where we could accurately predict all three behaviors and plotted their behavioral trajectories in Koopman space. First, we calculated a subspace where all three behaviors are separated based on the linear classifiers determined in the previous section. The first axis (Koopman subspace 1) was the found feature vector associated with digital screen usage. The second axis was the part of the socialization-associated feature vector that was orthogonal to the first axis. The third axis was the part of the physical manipulation-associated feature vector that was orthogonal to both the first and second axes. We then projected the Koopman state representation from the second (testing) annotated day onto these three axes and plotted trajectories when the participants were partaking in each behavior along with times where the participant was not doing any of the behaviors.

To test whether outwardly active behavior departed from the attractor state relative to wakeful rest (Figure 5A-right), for the nine participants with video annotations, we calculated the distance between each window's Koopman state representation and that participant's attractor state. We then averaged windows based on whether the participant was doing one of the three behaviors or awake but not doing any of the marked behaviors (which in practice meant sitting idly without obvious outward interactions). We used paired t-tests to ask whether the average active behavior to attractor distance was larger than the wakeful rest to attractor distance.

In Figure 5B, we assess the organization of neurocognitive states around this attractor. Specifically, we measure where neurocognitive states tended to form across participants along the axis between sleeping and actively waking states. For each participant, we calculated the mean Koopman state of all windows associated with active waking behavior and did the same separately for sleep. We then defined the vector between these two points as the participant's "sleep-wake axis" and projected the Koopman states from all windows during the annotated days of data onto this axis. Using a sleep score classifier (28), we classified windows associated with sleeping behavior into N1, N2, N3, and REM sleep. We then divided windows into the following groups: actively outward waking behavior, waking rest, N1, N2, N3, and REM. We calculated the mean of each group's projection along with the attractor state's projection onto this "sleep-wake" axis. To normalize the range of these projections between participants, we rescaled and recentered these projections linearly such that the center of all sleeping states was at zero and the center of active waking states was at one. We plotted the distribution of these state center projections in Figure 5B, and using paired t-tests with Bonferroni corrections, we assessed whether different states had reliably different locations along the sleep-wake axis across participants. We assessed whether different sleep stages activated different fMRI networks consistently across participants using the same method used to assess for consistent behavioral activation.

Supplementary Figures

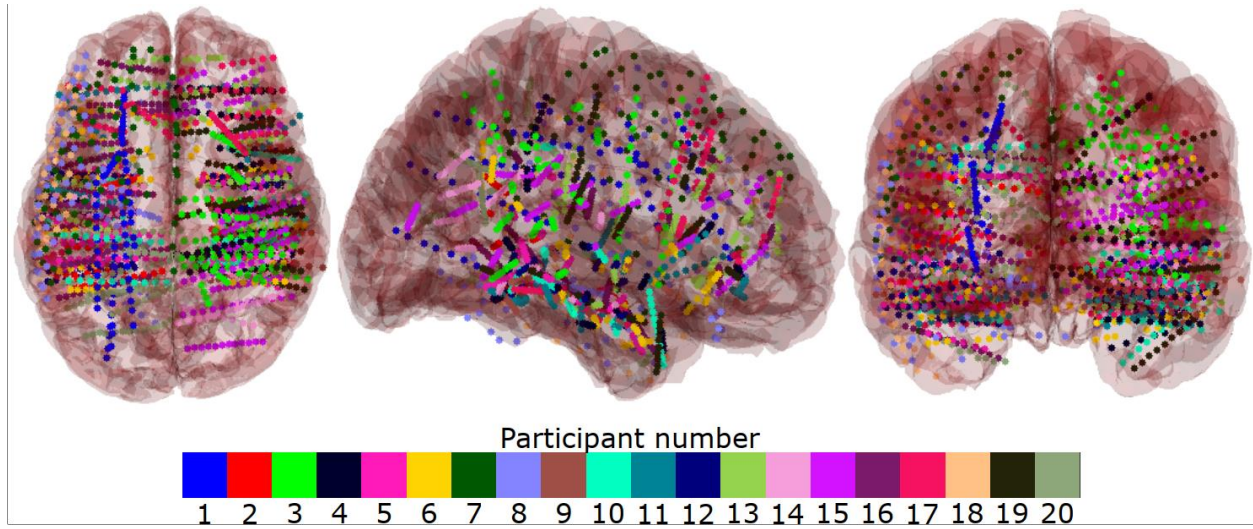


Figure S1: The location of each participant's electrodes in MNI coordinate space. Electrodes with the same color come from the same participant.

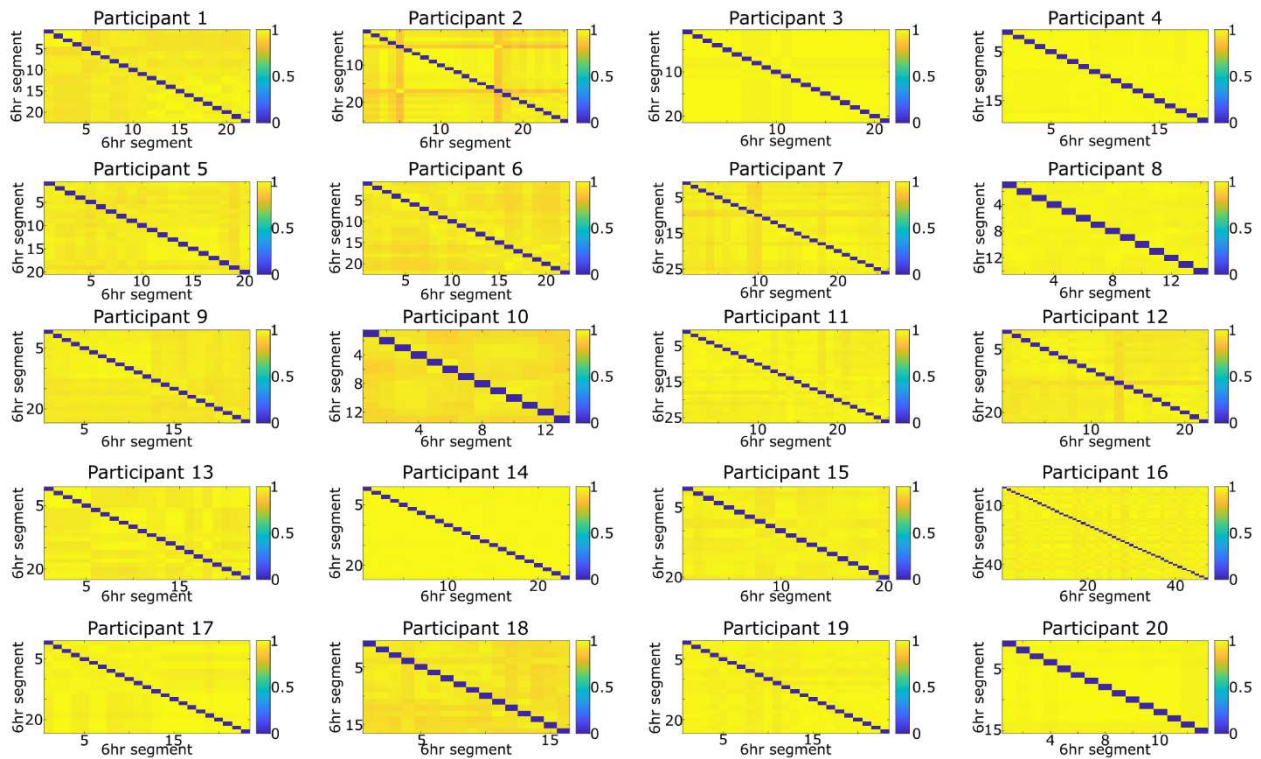


Figure S2: Parcellation stayed steady over time. We divided each participant's data into non-overlapping six-hour blocks and computed the optimal parcellation for each block. The Rand Index, which represents the proportional overlap between two parcellations of electrodes, is shown between all six-hour blocks for each participant. For each participant, we additionally calculated the average Rand index between all pairs of six-hour blocks. Across participants, this average value had a mean of 0.96 and standard deviation of 0.02, indicating parcellations remained highly consistent over time.

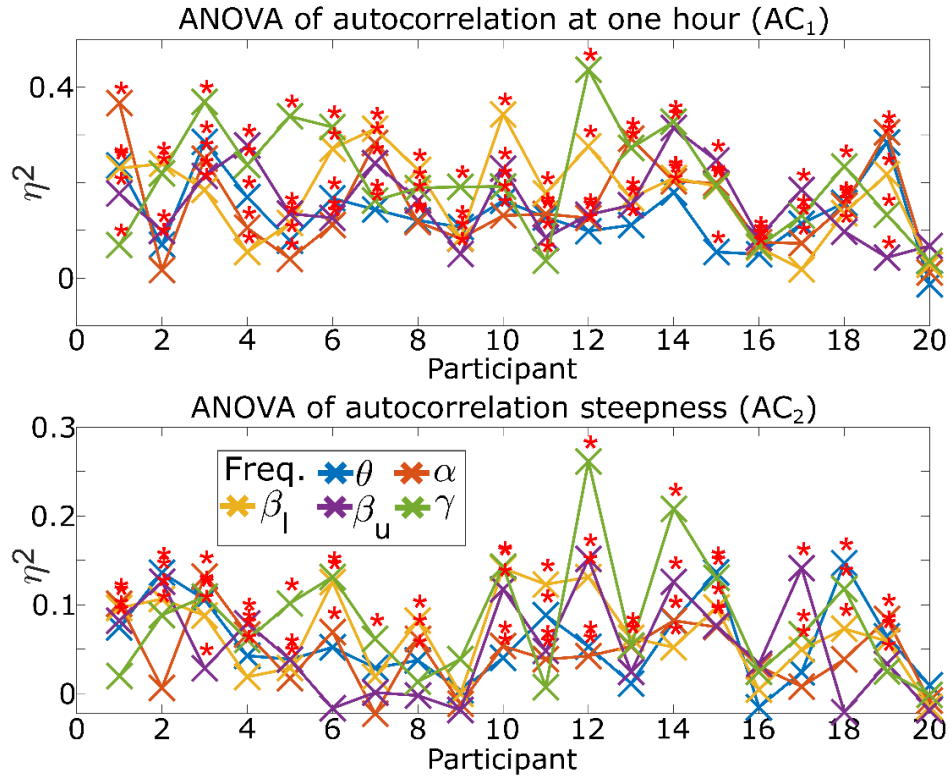


Figure S3: We assessed whether parcels had consistently different timescales using nonparametric ANOVA tests. For each participant individually, we divided their weeklong time-course into six-hour non-overlapping blocks. We calculated the autocorrelation of each parcel's coherence at a given frequency band (θ : theta, α : alpha, β_l : low beta, β_u : high beta, γ : gamma) across all blocks. We then tested whether the parcels from a single participant and frequency band had different autocorrelations from each other over these blocks using a Kruskal-Wallis one-way ANOVA test. Each group in the ANOVA test was the autocorrelations of a single parcel across all blocks, and we tested for whether there were differences in the group means. The effect size of the ANOVA test is shown above with asterisks marking statistically significant differences ($p < 0.05$). η^2 effect size indicates the percentage of variance in autocorrelation that is explained by which region autocorrelation was measured in.

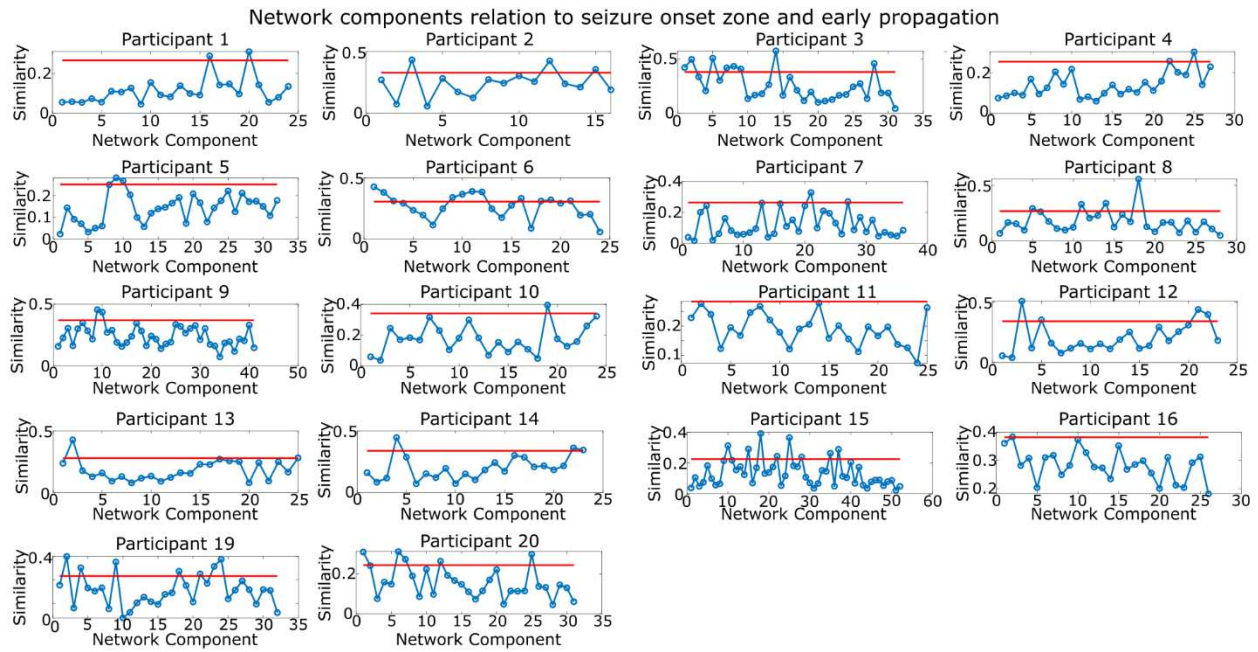


Figure S4: To ensure we remove all network components related to the participant's seizure related areas, we calculated the similarity between each participant's network component and their seizure zones and removed any network components showing above-chance similarity. Seizure zones were defined as any electrodes marked as part of the seizure onset zone or early propagation. Similarity was defined as the dot product between the absolute value of each participant's network component and their seizure zone and is shown above for all 20 participants. Participants 17 and 18 did not have any clinically defined seizure network. A null distribution for the dot product similarity generated by randomly permuting each network is shown with the red line to denote statistical significance threshold ($p=0.05$). All network components with significant similarity to seizure related regions were removed for all analyses on network components (Figure 2 and onwards).

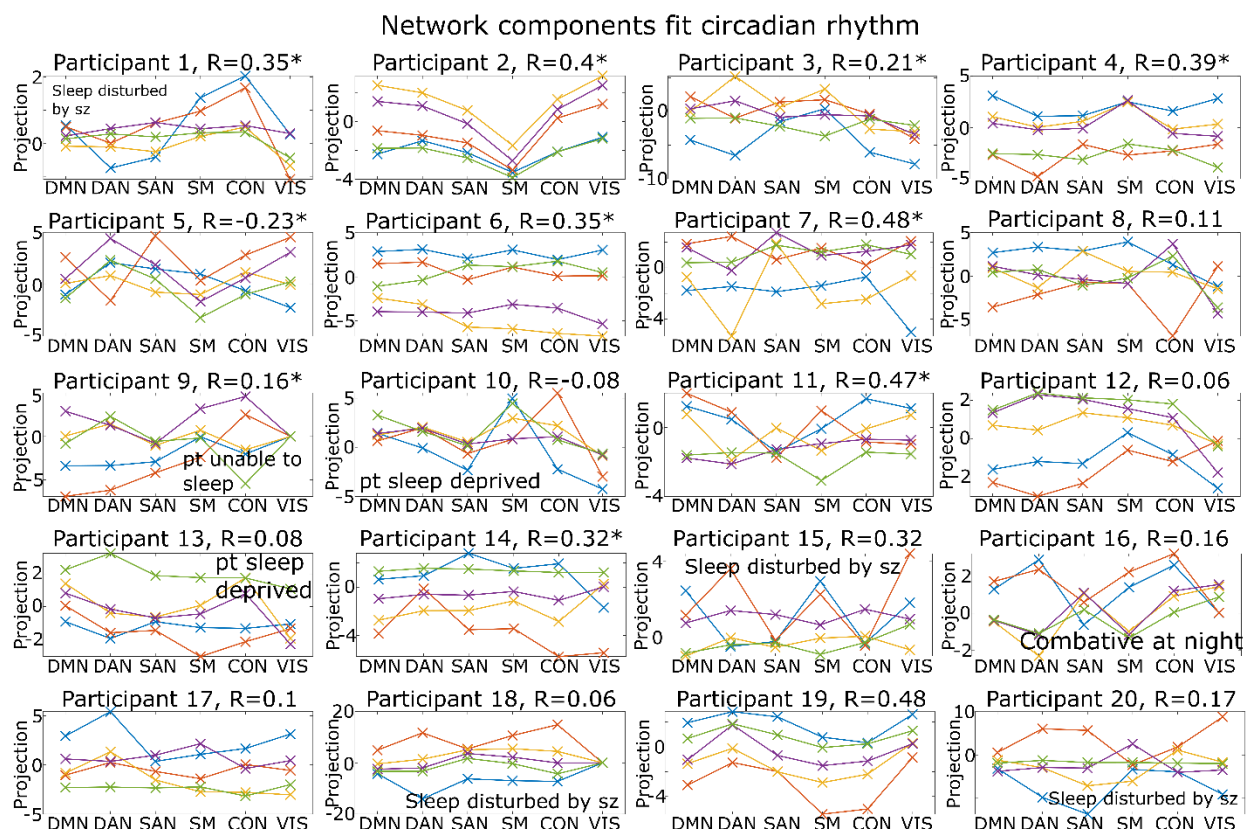


Figure S5: This figure shows the results in Figure 2A for all participants. The anatomical and frequency coverage of the mixture of networks associated with circadian rhythm in each participant are shown above. The correlation between a theoretical circadian sinusoid and the mixture's activation is shown above each plot (R). Asterisks indicate statistically significant correlations.

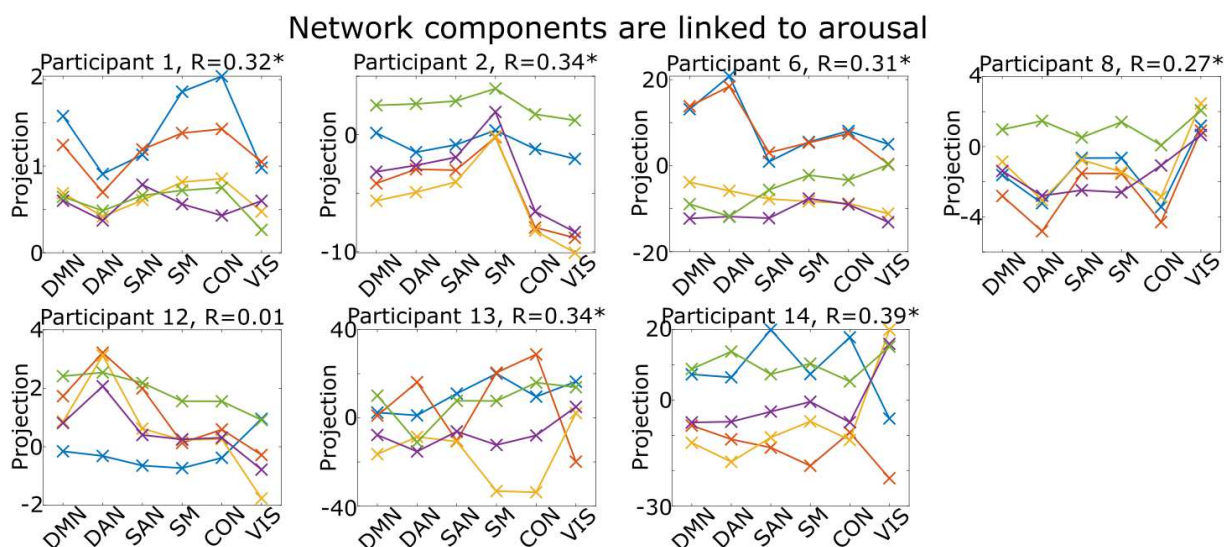


Figure S6: This figure shows the results in Figure 2B for all participants with recorded EKG. The anatomical and frequency coverage of the mixture of networks associated with heart rate (which is used as a proxy for arousal) in each participant are shown above. The correlation between the

heart rate and the mixture's activation is shown above each plot (R). Asterisks indicate statistically significant correlations.

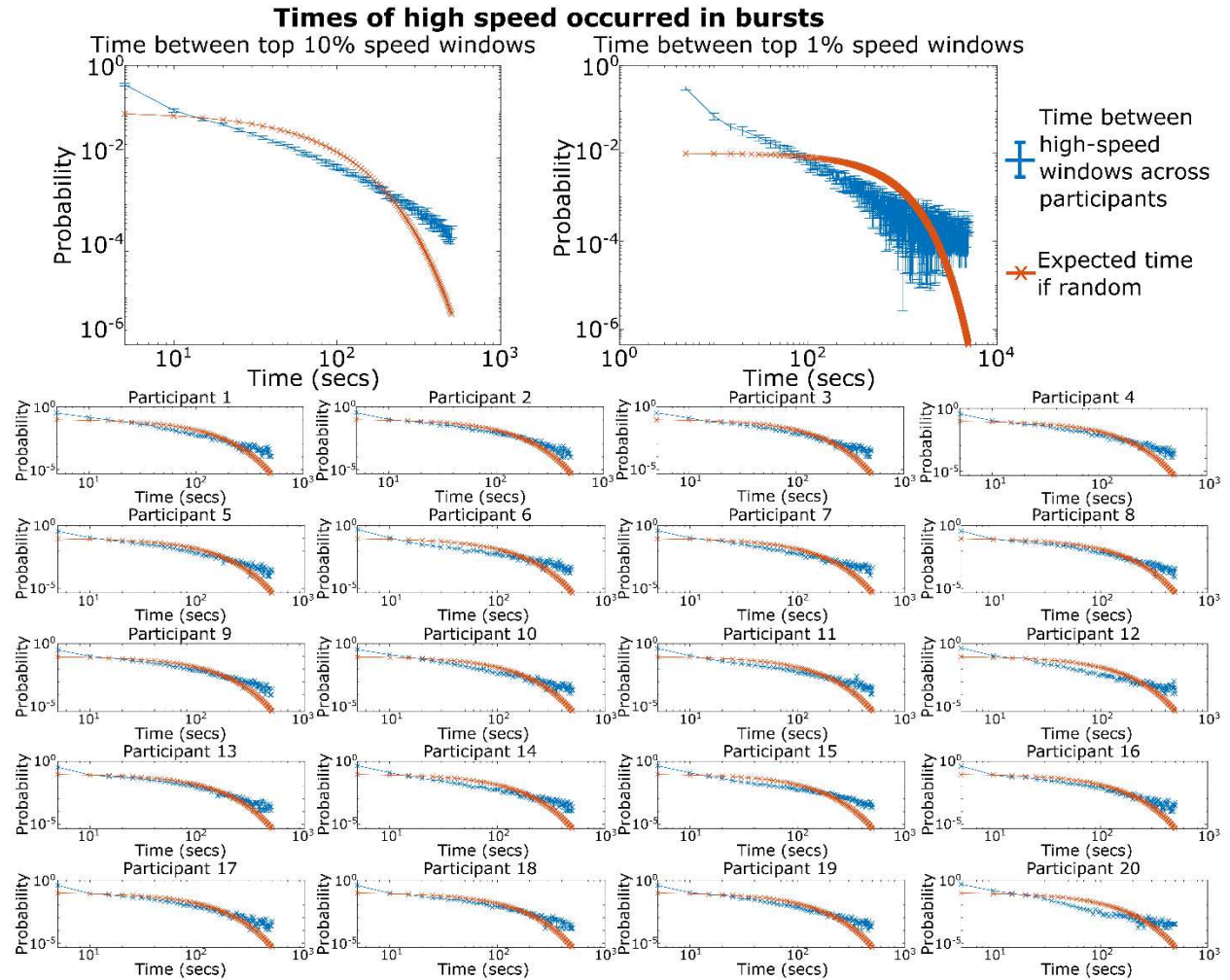


Figure S7. Top) The average time between windows with the top 1% or 10% of transition speed across all participants is shown in blue vs the expected time if windows of high speed occurred via homogenous Poisson process ($\lambda = 0.01, 0.1$). Error bars show 95% confidence intervals across participants. We found an increased occurrence of temporally adjacent or near-adjacent time windows of high transition speed, indicating that times of high speed occurred in “bursts” rather than in the distributed manner a Poisson process would indicate. We also found an increased occurrence in periods of time lasting for several minutes or longer with no times of high speed compared to what would be expected with a random Poisson process, indicating the existence of “stable states” of decreased speed. Bottom) The average time between windows of high speed for each individual participant. The distribution of each participant’s time between windows did not follow a Poisson distribution by Kolmogorov-Smirnov test ($p < 0.05$ in all cases).

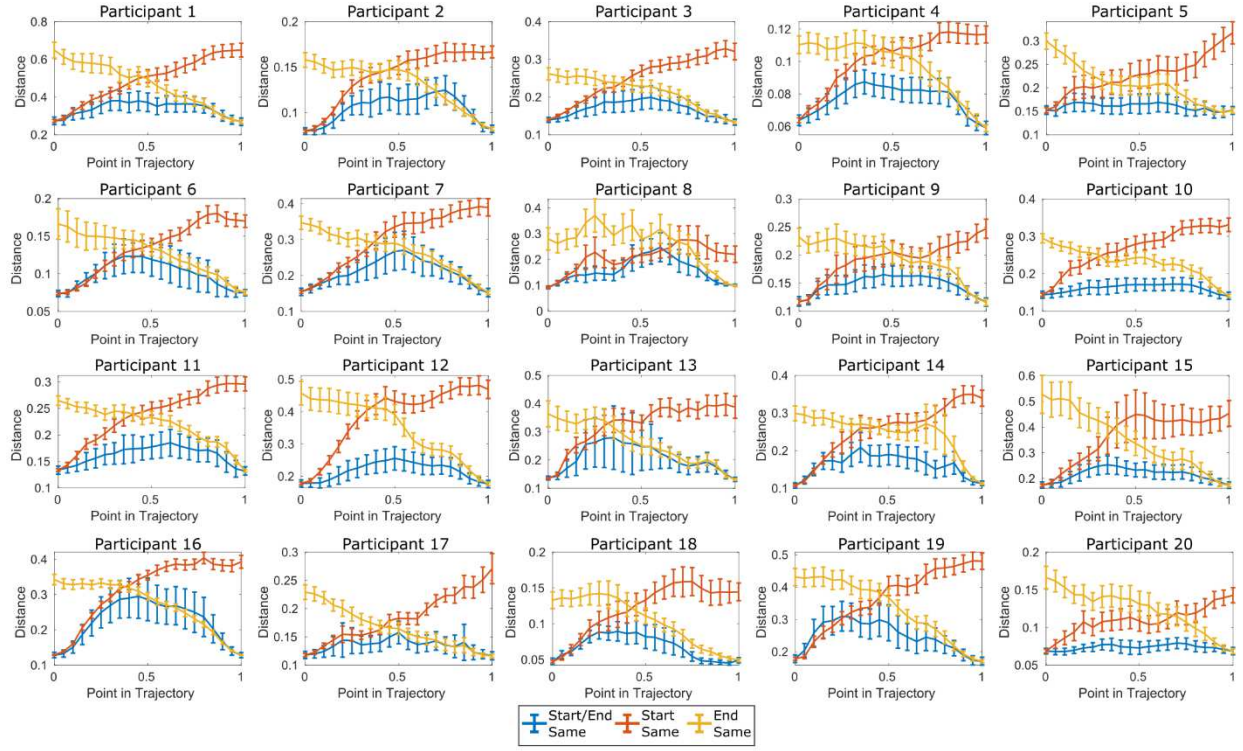


Figure S8: This figure shows the results in Figure 3D for all twenty participants. Average distance between pairs of transition trajectories as a function of what proportion of the trajectory was complete. Trajectory pairs are grouped into three categories: transitions with similar starting and ending points (1, blue) vs similar starting but different ending points (2, red) vs different starting but similar ending points (3, yellow).

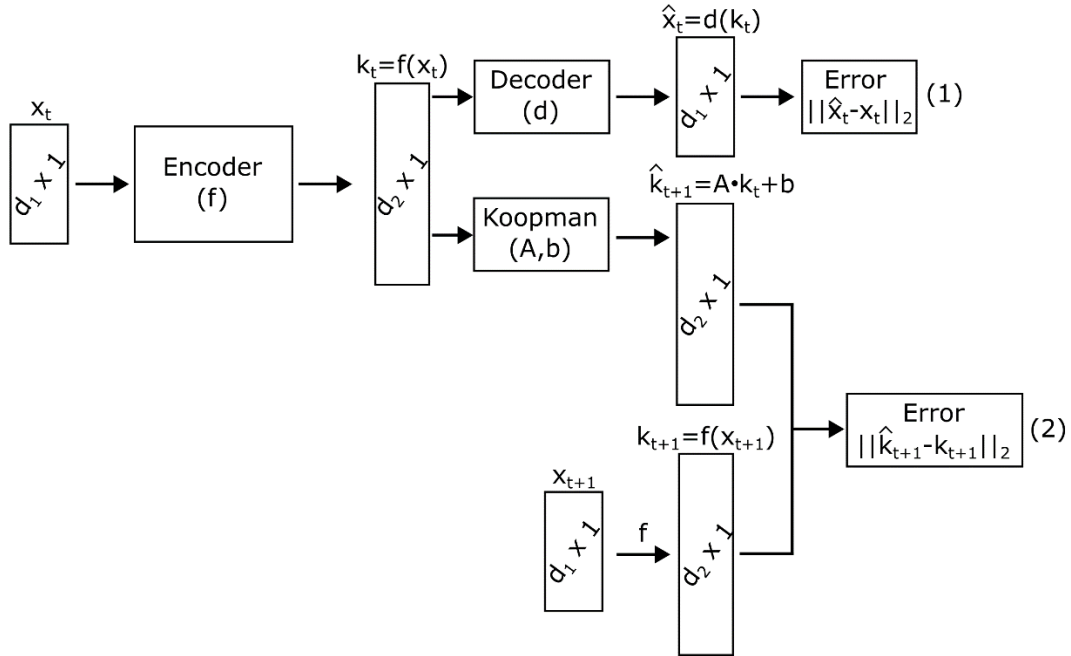


Figure S9: The Koopman model described in Figure 4A. x_t represents the network activations at time t . k_t is the output of the encoding model as is the Koopman state representation at time t .

Here the encoding model is a recurrent neural network, the decoding model is a standard neural network, and the Koopman operator (A,b) is a first-order discrete differential equation.

Algorithm 1 Training step

Feed the first half hour of $x_1, x_2, \dots, x_{30\text{min}}$ into the LSTM encoder f to initialize its internal state

Define $t = 30\text{min} + 1$ where x_t is the time window right after the initializing half hour

Define $k_{t-1} = f(x_{t-1})$, the output of the LSTM on the previous network activation state

Define the current error to be $\epsilon = 0$

while $t < \text{total number of time windows}$ **do**

$k_t \leftarrow f(x_t)$

$\hat{x}_t \leftarrow d(k_t)$

$\epsilon \leftarrow \epsilon + ||\hat{x}_t - x_t||_2^2$

$\hat{k}_t \leftarrow A \cdot k_{t-1} + b$

$\epsilon \leftarrow \epsilon + ||\hat{k}_t - k_t||_2^2$

$t \leftarrow t + 1$

end while

$\epsilon \leftarrow \epsilon/t + \text{L-1 regularization term of all learnable parameters}$

Update f, d, A, b according to the gradient of ϵ w.r.t learnable parameters

5

Figure S10: A single training step of the overall algorithm described in Figure S9.

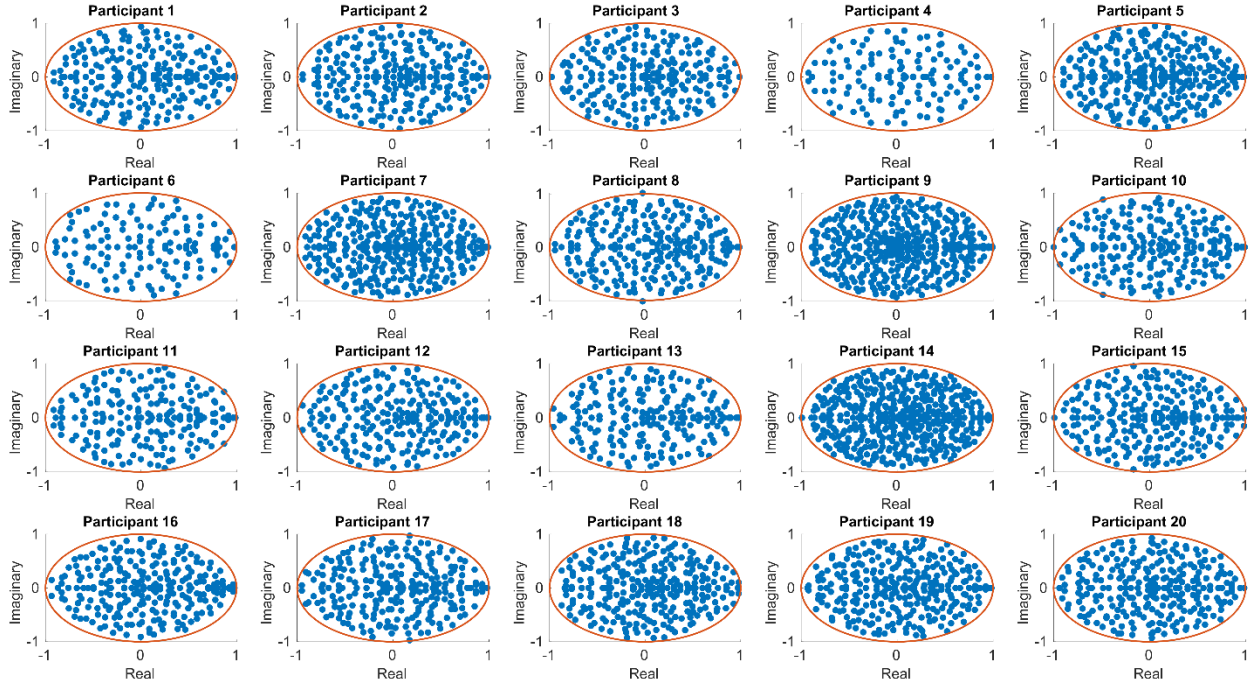


Figure S11: Eigenvalues of the Koopman operator (eigenvalues of the matrix A in Figure S10) over all twenty participants. In all participants, the real part of all eigenvalues is less than one, indicating that the overall dynamics were governed by stable attractors.

10

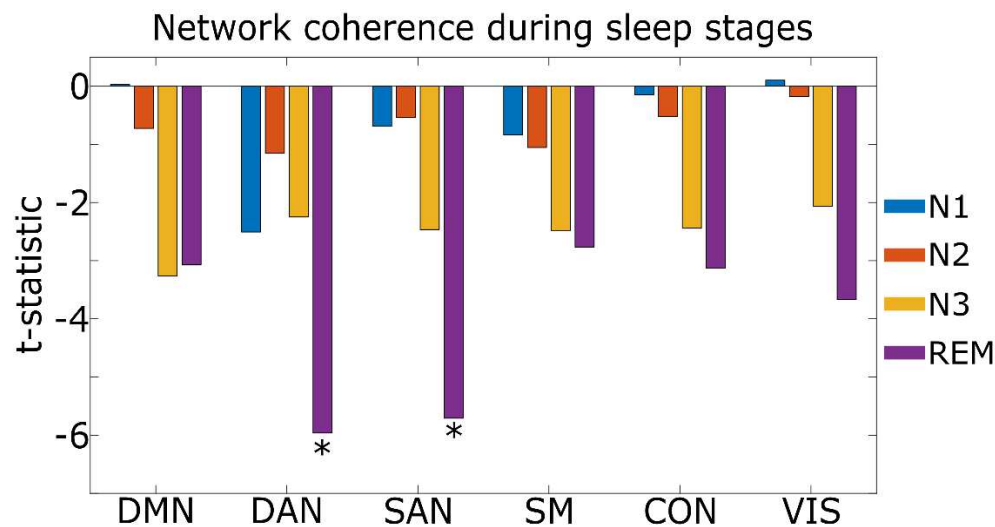


Figure S12: Anatomical regions consistently activated during different stages of sleep across participants. Negative t-statistics indicate that brain networks spanning the indicated fMRI network regions have lower coherence during the relevant sleep stage compared to other time windows throughout the week. Asterisks mark statistically significant networks post multiple comparisons.

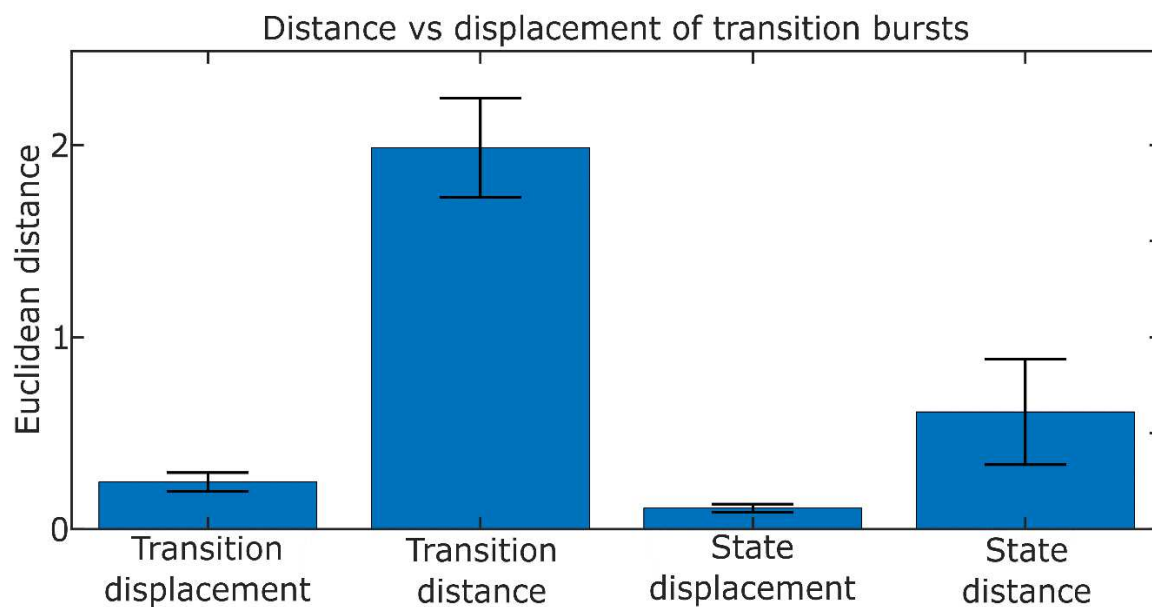


Figure S13: We tested whether brain transitory bursts went directly from one state to another or whether they took indirect, circuitous routes. We plotted the average total distance (sum of transition speed) traveled during transitions and stable states for all twenty participants versus their net displacement (distance between start and end states). Net distances during transitions were 8.87 ± 1.19 times larger than the net displacement (confidence interval indicates 95% bounds on the average multiplier across all participants). Net distance versus displacement during states were 5.99 ± 2.47 with the ratio between distance and displacement were higher for transitions than states ($p=0.01$, paired t-test).